DOCUMENT RESUME

ED 225 999 SP 021 828

AUTHOR Soar, Robert S.; Soar, Ruth, M.
TITLE Setting Variables, Classroom Interaction, and Multiple Pupil Outcomes. Final Report.
INSTITUTION Florida Univ., Gainesville. Inst. for Development of Human Resources.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C. Basic Skills Group.
PUB DATE Jun 78
GRANT NIE-G-76-0100
NOTE 119p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Academic Achievement; Anxiety; Classroom Techniques; Cognitive Style; Elementary Education; Formative Evaluation; *Individual Differences; Interaction Process Analysis; Locus of Control; Racial Differences; Sex Differences; Socioeconomic Status; *Student Characteristics; Student Motivation; Summative Evaluation; *Teacher Behavior; *Teacher Effectiveness; *Teaching Styles

ABSTRACT

Four general problems (two substantive, two methodological) were addressed in a research project: (1) Does the nature of the pupil or the setting make a difference in the teaching style which is most effective? (2) Does the cognitive level of the learning objective make a difference? (3) How can relationships within the classroom be analyzed? and (4) How should composite measures of classroom behavior be formed? Chapter one of this final report presents a brief overview of the project, which tested relationships between teaching style and pupil gain in a first and a fifth grade sample of pupils who differed in sex, race, socioeconomic status, IQ, anxiety, motivation, cognitive style, dependency, locus of control, or pre-test standing. Chapter two is a review of literature on research studies utilizing different methodologies: process-product relationship; trait-treatment interaction; and analysis at different levels of aggregation. The third chapter reports on a reanalysis of data collected from fifth and first grade classrooms in a previous project during 1969-70. Chapter four presents and discusses problems studied in this project: (1) to develop a set of process measures to serve as alternatives to the factor scores which had previously been used; (2) to develop procedures for analysis within classroom relationships; (3) to analyze interactions between characteristics of pupils and the teaching style which is most effective; and (4) to seek differences in the classroom behaviors which are associated with most effective learning for outcomes which differ in cognitive level. The fifth chapter presents a summary of the project and implications of the findings. (JD)

ED225999

FINAL REPORT

PROJECT NO. 6-0432
GRANT NO. NIE-G-76-0100

SETTING VARIABLES, CLASSROOM INTERACTION, AND MULTIPLE PUPIL OUTCOMES

ROBERT S. SOAR
FOUNDATIONS DEPARTMENT
INSTITUTE FOR DEVELOPMENT OF HUMAN RESOURCES
COLLEGE OF EDUCATION
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611
WITH
RUTH M. SOAR

JUNE, 1978

The research reported herein was performed pursuant to a grant with the National Institute of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government Sponsorship are encouraged to express freely their professional judgement in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

U.S. DEPARTMENT OF

HEALTH, EDUCATION AND WELFARE

NATIONAL INSTITUTE OF EDUCATION

BASIC SKILLS

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
✓ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

SP 021 828

Table of Contents

	Page
Acknowledgements	1
List of Tables	ii
List of Figures	iv
An Aid to Interpreting Tables	vi
1. Introduction	2
2. Review of Literature	4
Process-Product Relationships	4
Trait-Treatment Interaction	5
Analysis at Different Levels of Aggregation	5
3. Procedure	7
Pupil Measures (Fifth Grade)	
Achievement	7
Intelligence	7
Personality	7
Pupil Measures (First Grade)	8
Achievement	8
Intelligence	8
Cognitive Style	8
Socio-Economic Status	9
Classroom Observation Measures (Both Grades)	9
The Florida Climate and Control System (FLACCS)	9
Teacher Practices Observation Record (TPOR)	9
Florida Taxonomy of Cognitive Behavior (Cog Tax)	9
Reciprocal Category System (RCS)	10
Observational Data Collection Procedure	10
Analysis of Data	10
Calculation of Regressed Gain	10
Analysis of Process Measures	11
Analysis of Outcome Measures within Grade	14
Creation of a Pooled Data Set for First and Fifth Grades	15
Process-Product Analyses	16
4. Results and Discussion	21
Development of Process Measures	21
Logically Derived Subfactors	21
Factor Based on a Paradigm for Learning	23
Development of the Method of Analysis	37

Table of Contents - Continued

Page

Process-Product Analyses for First Grade 39
Screening 39
Models 42
Word Meaning 42
Listening 46
Matching 49
Numbers 49
Copying 49
Summary 54
Process-Product Analyses for Fifth Grade 57
Screening 57
Models 57
Vocabulary 57
Reading 65
Spelling 69
Arithmetic Concepts 69
Summary 73
Process-Product Analyses for First and Fifth Grade 79
Screening 79
Models 79
Complex 83
Simple 89
Total 89
Summary 93

5. Summary and Implications 94
Forming Composites for Classroom Process 94
Method of Analysis 95
Development of an Analysis for Within Class Relationships. . 95
Usefulness of Within Class Analysis 96
The Use of Models to Represent Process-Product Relations . . 97
Pupil Characteristics and Process-Product Relations 98
Pupil Characteristics and Gain 98
Process-Product Relationships 99
ifferences in Cognitive Level of Outcomes 100
A Concluding Comment 103

References 104
Appendix 108

List of Tables

Pages

4.1 Factor 1 - Negative Control vs Orderly Classroom 22

4.2 Subfactor Intercorrelations for Factor 1 22

4.3 Factor 2 - Teacher Central, Varied Interaction 24

4.4 Factor 5 - Free Movement and Positive Affect with Little Focus . 24

4.5 Correlations of Factors and Subfactors with Pupil Gain, Fifth Grade 25

4.6 RCS Items Added to Paradigm Factors 27

4.7 Factor Analyses of Items for First and Fifth Grade Combined Grouped by a Paradigm of the Classroom Environment for Learning-Emotional Climate 28

4.8 Factor Analyses of Items for First and Fifth Grade Combined Grouped by a Paradigm of the Classroom Environment for Learning-Management of Behavior 30

4.9 Factor Analyses of Items for First and Fifth Grade Combined Grouped by a Paradigm of the Classroom Environment for Learning-Management of Learning Tasks 32

4.10 Factor Analyses of Items for First and Fifth Grade Combined Grouped by a Paradigm of the Classroom Environment for Learning-Management of Thinking 34

4.11 Correlations Between Measures Based on a Paradigm of the Classroom Atmosphere for Learning, Grades 5 and 1 36

4.12 Comparison of Classroom-Level Effects for Vocabulary Gain Calculated from Classroom Means and from Individual Pupils, Fifth Grade 38

4.13 Comparison of Three Ways of Analyzing Within Class Relationships, First Grade 40

4.14 Variables Screened, Between Classrooms - First Grade 41

4.15 Variables Screened; Within Classroom - First Grade 43

4.16 Variance in Word Meaning Accounted for, First Grade 44

List of Tables (Cont'd) Pages

4.17 Variance in Listening Accounted for, First Grade 48

4.18 Variance in Matching Accounted for, First Grade 49

4.19 Variance in Numbers Accounted for, First Grade 50

4.20 Variance in Copying Accounted for, First Grade 51

4.21 Variance Accounted for by Preferred Models, First Grade . . . 53

4.22 Variables Screened, Between Classrooms, Fifth Grade 58

4.23 Variables Screened, Within Classrooms, Fifth Grade 59

4.24 Variance in Vocabulary Accounted for, Fifth Grade 60

4.25 Variance in Reading Accounted for, Fifth Grade 66

4.26 Variance in Spelling Accounted for, Fifth Grade 70

4.27 Variance in Arithmetic Concepts Accounted for, Fifth Grade . . 72

4.28 Summary of Variance Accounted for by Preferred Models, Fifth Grade 75

4.29 Summary of Variance Accounted for by First Order Models, Fifth Grade 76

4.30 Differences in Classroom Behavior From First to Fifth Grade 80

4.31 Variables Screened, Between Classrooms First and Fifth Grades 81

4.32 Variables Screened, Within Classrooms First and Fifth Grades 82

4.33 Variance in the Complex Outcome Accounted for, First and Fifth Grades 84

4.34 Variance in the Simple Outcome Accounted for, First and Fifth Grades 90

4.35 Variance in Total Outcome Accounted for, First and Fifth Grades 91

List of Figures

Page

3.1 A Paradigm of the Environment for Learning 13

4.1 Interaction of Differentiated Activities with Familiar Figures for Word Meaning Gain, Within Class, for First Grade 47

4.2 Interaction of Teacher Control and Negative Affect with Race for Listening, Within Class, for First Grade 50

4.3 Interaction of Differentiated Activities with Motivation for Vocabulary Gain, Between Class, for First Grade 62

4.4 Interaction of Differentiated Activities with Interest-Attention for Vocabulary Gain, Between Class, for Fifth Grade 63

4.5 Interaction of Teacher Gentle to Moderate Control with Pretest Reading for Vocabulary Gain, Between Class, for Fifth Grade 64

4.6 Interaction of Pupils Follow Routine with Externality for Reading Gain, Between Class, for Fifth Grade 67

4.7 Interaction of Teacher Gentle to Moderate Control with Interest-Attention for Reading Gain, Between Class, for Fifth Grade 68

4.8 Interaction of Pupils Follow Routine with Dependency for Spelling Gain, Within Class for Fifth Grade 71

4.9 Interaction of Narrow, One-Answer with Dependency for Arithmetic Concepts Gain, Within Class for Fifth Grade 74

4.10 Quadratic Relation Between Differentiated Activities and Complex Gain, Between Classes, First and Fifth Grades 85

4.11 Interaction of Teacher Strong Control and Negative Affect with Teacher Gentle to Moderate Control for Complex Gain, Between Class, for First and Fifth Grades 86

4.12 Interaction of Teacher Strong Control and Negative Affect with Grade for Complex Gain, Between Class, for First and Fifth Grades 87

List of Figures (Cont'd)

Page

4.13 Interaction of Pupil Negative Affect and Disorder with SES for Complex Gain, Within Class, for First and Fifth Grades 88

4.14 Interaction of Mid to High Cognitive Level with Pretest for Total Gain, Within Class, for First and Fifth Grades 92

Acknowledgments

In any project numbers of people smooth the way and help to make the product a better one than it otherwise would have been. At the National Institute of Education, Virginia Koehler and Joe Young have helped with advice and assistance with administrative details. Andy Porter, then a visiting scholar there, also made suggestions which were helpful, one of which would have saved us much time if we had understood it earlier.

Here at the University, Jim McClave of the Statistics Department led in the development of the method of analysis which has been employed, but of course, he is not responsible for its application.

In our Institute, Becky Montevideo and Kathy Poole have worked out administrative procedures and made the necessary arrangements with University Administration. Louise Stephenson has cheerfully and competently converted all our undecipherable hieroglyphics into clean copy and produced the result in the reader's hands.

To all of these we owe our thanks.

Robert S. Soar
Ruth M. Soar

June, 1978

Chapter 1
Introduction

Teacher effectiveness research which has related measures of classroom behavior with measures of pupil gain has established some relationships which appear to be dependable, but no one would argue that teacher effectiveness has been completely defined. Rather, there is a growing belief that we will learn more about teacher effectiveness if we take account of pupil characteristics or setting variables which make a difference in teaching effectiveness. The research step which follows from this belief is to identify pupils who differ from each other in certain preexisting characteristics, and test whether they differ in the style of teaching or classroom management which is most effective for them. Educators have always "known" that pupils differed in this respect, but there are surprisingly few results in the research literature.

This project was concerned with testing, in a fifth grade sample, whether the relationships between teaching style and pupil gain in achievement differed for pupils who differed in socioeconomic status, sex, race, IQ, anxiety, motivation, dependency, locus of control or pretest standing. In a first grade sample, parallel questions were asked for pupils who differed in sex, race, socioeconomic status, cognitive style, a rough measure of IQ and pretest. Since the first grade and fifth grade samples had the same classroom behavior measures available, this permitted testing whether pupils at different grade levels responded differently to the same classroom behavior.

In addition, some earlier results suggested that the style of classroom behavior which is most functional for pupil learning will also differ, depending on the cognitive level of the outcome measure. Past work suggests that a memory-level objective such as learning the multiplication table or dates in history is best achieved under rather closely focused classroom activity, but that if complex problem-solving or creative production are the objective, more freedom for pupils to explore ideas and less closely focused interaction are useful. That question will be examined.

A related issue which emerged during analysis was the effect of level of aggregation of the data on the results obtained. It was also examined. For years, sociologists and economists have expressed strong concern for this issue, recognizing that the relationship between two variables may change from one level of aggregation to another, and possibly even change in direction. Past work in teacher effectiveness, on the other hand, has usually been limited to the study of relationships between classroom means, that is, between overall measures of classroom behavior and mean pupil outcomes. In the few instances in which pupil characteristics have been studied in relation to teaching effectiveness, classroom means have also been the unit of analysis, because of the methodological constraint that data from individual pupils in a classroom are not independent, and because of lack of knowledge of an acceptable alternative. Recent developments in analysis have suggested a way

of dealing with this difficulty and a method was evolved for use in this project to examine both between class and within class relationships.

A final issue which was examined was that of exploring alternative ways of forming classroom behavior scores. Individual items seemed undesirable because they were so numerous, to a degree redundant, and typically low in reliability. But the factor scores used in previous work were sometimes conceptually complex and difficult to interpret. Two other alternatives were examined.

In summary, four general problems were addressed--two substantive, and two methodological: Does the nature of the pupil or the setting make a difference in the teaching style which is most effective? Does the cognitive level of the learning objective make a difference? How can relationships within classroom be analyzed? and How should composite measures of classroom behavior be formed?

Chapter 2
Review of Literature

Three separate literatures are relevant to this report: process-product studies which have looked at relationships between classroom behavior and pupil learning and which have not considered pupil characteristics; studies of trait-treatment interaction, which have studied the interaction between pupils and teaching styles; and finally, the literature which deals with analysis of data at different levels of aggregation, and specifically, analysis of data both between classrooms and within classrooms.

Process-Product Relationships

These studies have been reviewed recently, extensively and competently, so only brief mention will be made of them. The most recent review is that by Medley (1977) which does an admirable job of organizing the results of previous research. Findings are organized by the socioeconomic status of the pupil group; by whether the outcome measure is reading, arithmetic or a noncognitive measure; by whether the outcome measure is at a low or high cognitive level; and finally by whether the pupil group is at the primary or intermediate grade level. As a consequence of this organization, some results which have appeared inconsistent in other reviews are quite consistent. For example, some teacher behaviors relate relatively consistently in one direction for high SES pupils, but in the other direction for low SES pupils. There are similar differences for the lower and higher grade levels. Of course, this is a beginning toward the identification of the different teaching behaviors which are functional for different pupil groups, except that entire studies are used as the data source rather than pupils within a study.

The most extensive recent review is that of Dunkin and Biddle (1974), which reviews presage-product, process-product, and presage-process studies. Earlier reviews are those of Rosenshine (1971), and of Rosenshine and Furst (1975).

Rosenshine and Berliner (1978) have summarized recent work on classroom behavior and pupil achievement gain and concluded that achievement gain is related to such variables as amount of material covered, amount of time engaged in academic work, teacher direction of pupil activity, teaching large groups, teaching by the use of low cognitive level questions, and the absence of negative affect. They characterised this pattern of behaviors as "direct instruction." They did not suggest differences in recommended teaching style for different pupil groups or for different objectives.

In contrast to that point of view, an integration of the results of four studies done by the authors (Soar & Soar, 1976) indicates a degree of consistency in finding that for higher cognitive level objectives a greater degree of freedom of choice and less focus in the learning task was functional than was true for lower cognitive level objectives. There is limited data to support this in the Medley review cited earlier, and also in Stallings (1975), in which there was greater growth for the Ravens Progressive Matrices for Follow-Through programs which were less structured; in contrast, there was greater growth in narrower learning objectives for more highly structured programs.

Trait-Treatment Interaction

The major review in this area is Cronbach & Snow (1977). However, only one study was from natural classrooms, and as the authors comment, that study was improperly analyzed, using pupils as the unit of analysis, and not recognizing that they were organized into classrooms. All of the other studies were experimental studies.

The two studies that have examined trait-treatment interactions in natural classrooms are those of Soar & Soar (1973), and Brophy and Evertson (1974). In the Soar & Soar study, pupil characteristics as well as outcomes were aggregated to the level of classroom means and interactions between mean pupil characteristics and classroom process were tested. In addition, the data of individual pupils were analyzed and tabled, but the problem of nonindependence of observations within classroom was recognized, and no conclusions were drawn. In the Brophy/Evertson study, data were aggregated to the level of classroom means and relationships between classroom process and pupil mean gain were studied across two sets of schools--Title I and non-Title I--with differences in relationship which were consistent over two years of data interpreted as interactions of process with pupil socioeconomic status. (However, Brophy has later pointed out that SES is a proxy measure for other characteristics such as ability, achievement level and motivation (Medley, 1977).

Analysis at Different Levels of Aggregation

The fact that a relationship between two variables may differ when that relationship is calculated at different levels of aggregation has been a concern for some time, particularly for sociologists and economists (Dogan & Rokkan, 1969). A problem for these disciplines has been that data typically come to them aggregated by census tract, county, state, region, or nation, but the relationship which is sought is sometimes that between individuals. Methodological articles make clear that the occurrence of differences in relationship at different levels of aggregation is well understood in those fields. The data in teacher effectiveness research come both at the level of the individual pupil and at the level of the classroom, but methodologists in our field have argued that the classroom is the only appropriate level of aggregation because of the interdependence of pupils within classroom (Raths, 1967; Wiley & Bock, 1967; Peckham, Glass & Hopkins, 1969; Glass & Stanley, 1970). As examples of this interdependence, Peckham et al., point out that a behavior problem in a classroom may hamper learning for all pupils in that classroom, or that the child who asks good questions may support learning for all other pupils. These authors have ignored the possibility that considerable portions of the variance in a given pupil characteristic might exist within classroom, rather than between classrooms, and the possibility that relations might differ from one level of aggregation to another. In support of the latter point, Hines (n.d.) suggests that the relationship between mental age and chronological age within grade level will probably be a moderate negative value, since the older pupils in a grade are likely to be those who have been held back, and consequently those who are less able.

But the relationship between mean MA and mean CA across grade levels will be positive and strong. That is, the relation for individuals within grade will be negative, whereas the relationship between grades will be positive.

When the question is whether pupils of different characteristics may differ in response to a particular kind of teaching, it seems particularly limiting to ignore within classroom variance. It seems reasonable that pupils would be less likely to be grouped on the basis of traits such as anxiety, dependency or motivation than on the basis of intellectual characteristics, so it seems probable that interactions between noncognitive variables and classroom process might be stronger within groups than between groups.

A resolution of this methodological impasse has been suggested by Cronbach and Snow (1977), in which the recommended procedure was sketched only briefly. Cronbach (1976) (cited in RIE in the summer of 1977) spelled out in careful detail the need for this sort of analysis, the circumstances under which relationships might be strong within class but not between classes, vice versa, or both, and concluded that most educational research data collected from pupils who are members of classrooms have been improperly analyzed. As examples, he suggested that comparisons of curricula, item analyses and factor analyses ought to be done at least at these two levels of aggregation. He raised the possibility of aggregating to school, system, or still higher levels, and suggested that choice of aggregation level or levels is an important decision in any research.

Ironically, there is an example of this recommended procedure in the literature nearly 20 years ago. Medley and Mitzel (1959) analyzed pooled relations between classroom behavior and pupil gain within school, after first partialling out between school relations. It is unfortunate that their solution to the problem of multiple levels of analysis was not immediately recognized and used by researchers.

A major part of this project, then, was the study of trait-treatment interactions of the personological characteristics of pupils and grade levels with measures of classroom behavior, analyzed both at the within and between classroom levels.

Chapter 3
Procedure

This report is a reanalysis of data collected in fifth and first grades in a previous project during the period 1969-71.[1] The instruments and data collection procedures are described in detail in the report of that project (Soar and Soar, 1973). Both cognitive and non-cognitive measures were collected from pupils in the fall, classroom observation was carried out in the middle of the year, and pupils were retested in the spring.

Pupil Measures (Fifth Grade)

Achievement

Pupil achievement was measured by the Reading, Vocabulary, Spelling, and Arithmetic Concepts subtests of the Iowa Tests of Basic Skills (ITBS) (Lindquist and Hieronymus, 1956). These subtests were selected as skills which are particularly important to success in school and to competence in the larger society, as representing both verbal and quantitative skills and as representing a range of measures from abstract to concrete.

Intelligence

The Verbal subtest of the Lorge-Thorndike Intelligence Tests (LT) was used. It has adequate reliability for use with individuals, and correlations both with individual intelligence tests and with pupil achievement indicate its validity as a measure of school learning potential (Freeman, 1959).

Personality

Dependence-Proneness Scale (DP). Although the original research (Flanders, Anderson, and Amidon, 1961) was done with eighth-grade pupils, it appeared to present no difficulty for elementary pupils in grades 3-6 in a previous project (Soar, 1966). The validity of the scale was supported by a number of studies relating scores to observed classroom behavior in the original research.

Children's Manifest Anxiety Scale (CMAS). This scale (Castaneda, McCandless, and Palermo, 1956) was standardized on fourth-, fifth-, and sixth-grade children. It has adequate reliability, and has demonstrated predicted relationships with simple and complex learning in studies done by its authors (Castaneda, Palermo, and McCandless, 1956).

---

[1]Performed under grants numbered 5 ROI MH 15891 and 5 ROI MH 15626 from the National Institute of Mental Health, Public Health Service, Dept. of HEW, Washington, D.C.

Junior Index of Motivation (JIM). This scale has been shown to differentiate underachievers from overachievers with intelligence partialled out; to relate to teacher judgments of pupil motivation; and to correlate with raw achievement measures, +.37 to +.40 (Frymier, 1965). The relation of the items to motivation is not generally obvious, with items like: "It is better to forget than to forgive." "There is a real limit to man's intelligence." "The dreamer is a danger to society." "Experience may be a good teacher, but schools are better."

Social Reaction Inventory (SRI). The scale is a modification of Rotter's (1966) I-E Scale, a measure of internality or externality of perceived control. The modification was intended to simplify the reading level to make it meaningful for elementary aged children. Its use was frankly experimental.

Socio-Economic Status (SES). Socio-economic status was assessed using the occupation of the family breadwinner (ordinarily the father, but the mother if she supported the family), taken from school records. The classification scheme followed was that of Warner, Meeker, and Eells (1949). Duncan (1961) has shown that the loss in precision in estimating social status using occupation alone is small, in comparison with using all sources of information available in census data.

Pupil Measures (First Grade)

Achievement

The Metropolitan Readiness Tests (MRT) (Hildreth, Griffiths, and McGauvran, 1969) were used. This is a widely used battery whose reliabilities and predictive and content validities are all satisfactory.

Intelligence

The Draw-A-Man test from the MRT was used as a rough estimate of intelligence.

Cognitive Style

The Matching Familiar Figures Test (MFF) is an instrument developed by Kagan (1965) in which each item presents the child with a stimulus form and six alternative response forms which are similar but only one of which is identical to the stimulus. The measure is assumed to assess a reflective versus impulsive style--the willingness to continue effort in an ambiguous situation rather than dismissing the task with a perfunctory response.

The test was designed for individual administration with both response latencies and numbers of errors recorded, but this procedure was impractical with the numbers of pupils from whom data were collected here, so the instrument was given typical group administration. The change in procedure may have invalidated the test, of course. But since the "error measure" was the most discriminating in the development

of the instrument (Kagan, 1965), it was included on an experimental basis. Pupils did appear to take it with interest; so that it could even be used to "lighten" a testing period.

Kagan (1965) and Kagan, Pearson, and Welch (1966) indicated that the test was related to reading skill and conceptual reasoning among first grade pupils with WISC verbal ability controlled statistically.

Socio-Economic Status

The same measure was used as in the fifth grade data.

Classroom Observation Measures (Both Grades)

The Florida Climate and Control System (FLACCS)

This was a modification of the South Carolina Observation Record (Soar, 1966), which drew heavily on the Hostility-Affection Schedule (Fowler, 1962), and the earlier versions of the Observation Schedule and Record (Medley and Mitzel, 1958, n.d.). The first section is organized around the concept of direction and control of the classroom by the teacher and the response of pupils. Twenty items of verbal teacher control are scaled into five levels of coerciveness ranging from "Guides, Suggests," to "Orders, Commands". A smaller number of nonverbal items are similarly scaled. The pupil items represent response to teacher control such as assumption of responsibility for classroom order, pupil task involvement, obeying, resisting or disobeying, and the organization of the classroom. The second section records teacher and pupil affect expression, verbal and nonverbal, positive and negative.

Teacher Practices Observation Record (TPOR)

In this instrument, every item describes a teacher behavior that is widely practiced in schools, but half the items describe behavior which reflects agreement with Experimentalism and would be espoused by John Dewey; the other half reflect disagreement (Brown, 1968). Brown has done extensive research with the instrument relating it to measures of beliefs, and has shown relations between a teacher's beliefs and teaching practices.

Florida Taxonomy of Cognitive Behavior (Cog Tax)

The original instrument, which was used in the fifth grade sample, was developed by the Florida group under Brown's leadership. Its history originates with The Taxonomy of Educational Objectives: The Cognitive Domain (Bloom and others, 1956), which was modified and extended by Sanders (1966) to provide an instrument to assess teacher lesson plans and teaching materials. The work of the Florida group consisted of converting Sanders's instrument to one for live observation in the classroom, and of carrying out developmental work with it in upper grade classrooms.

This version was found difficult to apply in kindergarten-first grade classrooms in earlier research, and a new version of the instrument was developed. It was used in first grades in this study. Since the items in the two instruments were different for the two grades, only the sums for each cognitive level, prorated for numbers of items, were used.

Reciprocal Category System (RCS)

This instrument is a modification of the Flanders system; work with the original system has been summarized by Flanders (1965, 1970), and Amidon and Hough (1967).

Ober, Wood, and Roberts (1968) expanded the seven teacher categories of the Flanders System to nine: teacher lecture was divided into that which is responsive to pupils and that which is teacher initiated, and the category of teacher criticism was divided into one for correction without criticism, and one for criticism. Each of the categories was redefined so that it could be used for pupil talk as well as teacher talk.

Factors from these four instruments discriminated between grade levels and programs in Follow-Through, and were related to pupil gain (Soar and Soar, 1972; Soar, 1973).

Observational Data Collection Procedure

Data collection was carried out by teams of two observers, each of whom was trained in both TPOR and FLACCS. Each observer used each instrument for approximately half of his classroom visits, so that any biases remaining after training would be minimized by having the total group of observers use both instruments. Observers were varied across teachers for the same reason. Finally, teachers were observed on different days of the week and at different hours of the day to minimize effects associated with these variables.

Three five-minute periods of observation on a single occasion completed a series, and eight series were sought for each teacher, although only six or seven were collected in some cases. Simultaneously with each observation the observers also made a tape recording, which was later coded on the RCS and Cognitive Taxonomy by separate teams of coders. As a consequence of occasional technical difficulties, the number of tape recordings for a teacher was sometimes smaller than the number of times a team observed in the classroom, reducing the data available for some teachers for the RCS and Cog Tax.

For all four instruments, data were reduced to means per series.

Analysis of Data

Calculation of Regressed Gain

In the previous project, regressed gain was calculated for subgroups

of pupils identified by sex and race at both grade levels, additionally by narrow ranges of I.Q. in the fifth grade, and by socioeconomic status level in the first grade. This presumably gained the advantage identified by McNemar (1975) that the use of at least two covariates minimizes the overadjustment-underadjustment dilemma (Campbell and Erlebacher, 1970; Bereiter, 1963). The procedure typically accounted for from 80 to 90 percent of the variance in postscore, in contrast to the more common regressed gain procedure in which only pretest is held constant, typically accounting for approximately two-thirds of the postscore variance. While this lessens the reliable variance which is available for further analysis, it probably also lessens or eliminates a bias which is potentially important. The previous project found moderately strong relationships between pretest characteristics of pupils and teacher behavior at midyear. If pretest characteristics of pupils are not partialled out of the gain measure adequately, then those characteristics will relate both to outcome and to teacher behavior, producing spurious relationships between them. For example, if teachers with high IQ classes encourage more pupil initiation, and if IQ is related to gain, then pupil initiation will appear to be related to gain, but the relation will be spurious. The method for calculating gain used in this study should have minimized that possible bias if not eliminated it; it also fitted different regression slopes between pre- and posttests for different subgroups, need for which was apparent in the data.

Analysis of Process Measures

In the original study, the classroom observation data were reduced by first factoring each of the four observation instruments separately, carrying forward the high loading items, and refactoring that subset of items. Difficulties of interpretation were sometimes encountered, however, since the empirical procedure sometimes put together subsets of items which were not clearly conceptually related, such as positive affect and a high level of physical activity.

Several alternative procedures were considered. At the outset, the use of individual items of the observation systems was rejected, primarily because a large number of relationships would be tested, but also because of the low reliability of individual observation items, and because teacher behaviors do not occur in isolation--rather, they occur in clusters which support or moderate each other. As a consequence, some procedure for creating composites or clusters of items seemed necessary.

Factoring valid items. Another procedure which was considered was that of relating individual items to pupil outcome measures, retaining only those which showed some minimum level of relationship, and factoring that subset of items. This procedure was not followed because we were reluctant to begin the analysis by selecting items which had some minimum level of validity. Not only did this leave us with a concern for biasing later tests of validity, but also with concern that teaching probably takes place by means of clusters of behaviors which teachers

employ in constellations, but some items in the cluster might not be related to pupil gain. Retaining only valid items would then destroy the constellations and risk altering the meaning of the items.

Logical Analysis of Factors. Where the original factors appeared to be made up of subsets of items which were not clearly logically related, a trial procedure was followed in which logically coherent subsets of items were identified, examined for internal consistency, subfactor scores were calculated, and the relationships between these subfactors were calculated. Where difficulties of interpretation in the original multiple regressions using factors had occurred, reanalyses were carried out using the logical subfactors.

Factored Rational Item Subsets. In the course of thinking through process-product relationships from four earlier sets of data (Soar, 1966 and 1968; Soar & Soar, 1972; Soar & Soar, 1973) a paradigm of the environment for learning which seemed to fit those past findings emerged. Figure 3.1 from Soar & Soar (in press) presents that rational organization.

The major headings, Emotional Climate and Teacher Management (or control) are parallel to the factors which emerged from early studies of leadership of adult work groups (Fleishman, 1953), and which were found to be relatively independent dimensions of teacher behavior (Soar, 1966). However, Emotional Climate is not often separated from Management in educator's thinking, nor are the three areas of Management of Behavior, of Learning Tasks, and of Thinking Processes distinguished from each other. But there is some evidence that these are distinctions which may be useful in attempting to identify effective teaching.

As an illustration, Emotional Climate and Teacher Management are combined in the traditional concept of permissiveness, which suggests that the permissive teacher is warm and supportive and also shares decision-making with pupils. The nonpermissive teacher is seen as negative and critical and directs the activities of the classroom firmly. However, our early work (Soar, 1966) indicated that classrooms with a positive emotional tone may either be tightly organized and teacher-directed (as the typical contingency-management classroom is), or may have little organization or task direction. The negative classroom may be one in which the teacher uses negative affect and criticism of pupils to "run a taught ship", or it may be the occasional classroom where the teacher spends the day screaming at pupils but never establishes enough order to teach. Thus, it seems important to distinguish emotional climate from management and control.

Teacher Management consists of three areas of classroom activity. Pupil Behavior refers to such things as the amount of physical movement, frequency of pupil socializing and fluidity of groupings--the non-task, behavioral aspects of the classroom. Management of Learning Tasks refers to the way learning tasks are selected and carried out in the classroom--the degree of control exercised by the teacher or the amount

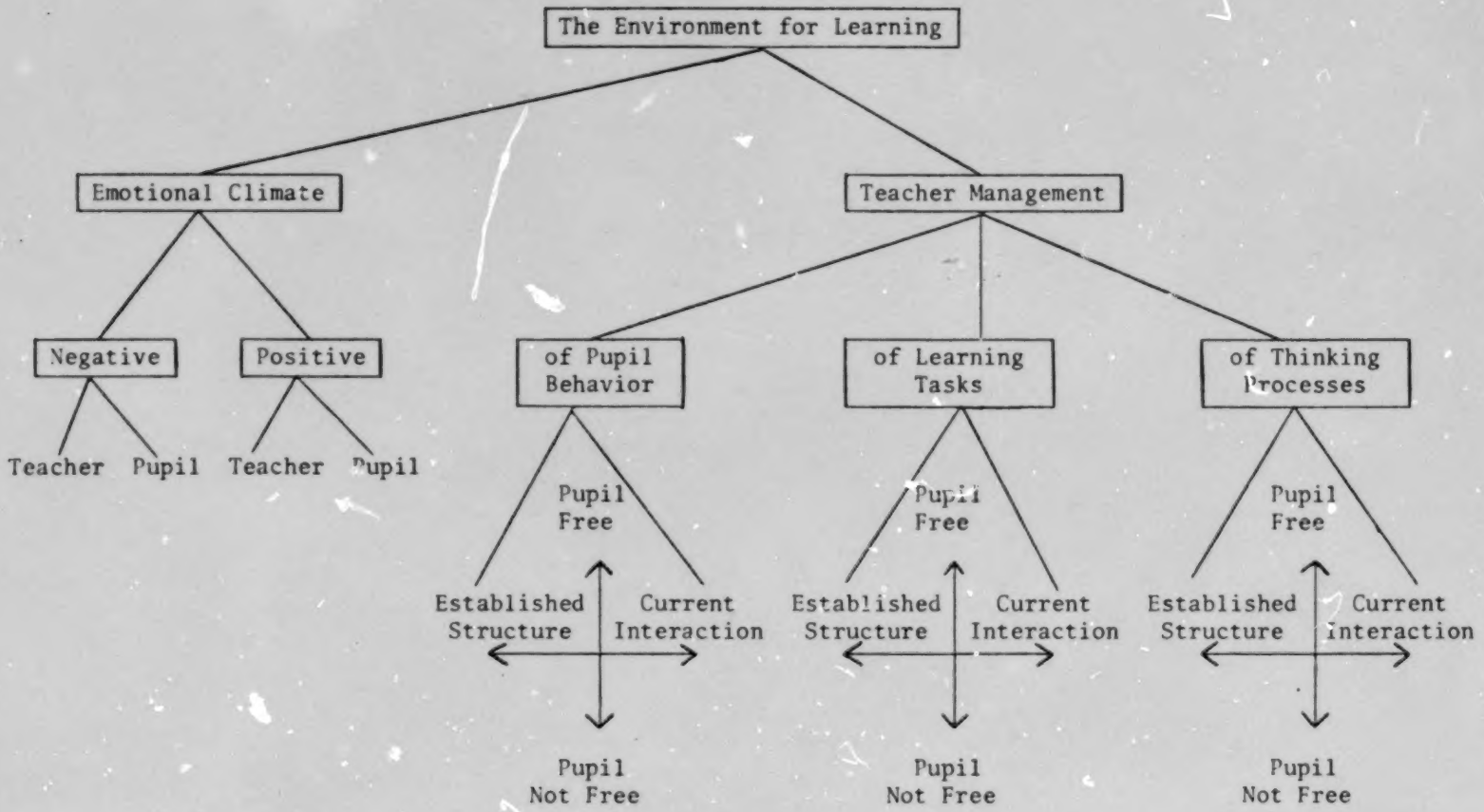


Figure 3.1

A Paradigm of the Environment for Learning

of choice permitted pupils. Thinking Processes refers to the manner in which thought is developed in the classroom--the freedom (or lack of freedom) the pupil has to explore related ideas, or even divergent ones, and the cognitive level of the thinking processes permitted or encouraged by the teacher. These three areas of teacher management are not often separated. As extreme examples, contingency management classrooms (which use conditioning principles and programmed learning) typically control all three areas closely. Open classrooms, on the other hand, attempt to free thinking processes and learning tasks, but in doing so, usually free behavior as well--sometimes to the point of permitting chaos.

The continuum whose ends are identified as Established Structure and Current Interaction under each of the management subheads was not utilized in the development of measures. Measures to represent this rational scheme were created by first selecting items on an a priori basis from three of the four observation instruments (omitting RCS) to represent each domain. After pools of items representing the domains were selected on a rational basis, each pool was factored, and the factor analyses were used to create unit-weighted incomplete factor scores.

The attempt to reproduce the paradigm was not completely successful, although the major elements survived. The resulting factors are reported in the Results chapter.

Items from the Reciprocal Category System were treated differently, since it was often not clear whether an item referred to management of behavior, subject matter or of thought. As a consequence, after factor scores had been developed representing each domain of the paradigm from the other three instruments, those factors were related to the RCS items and the RCS item was added to the factor with which it was most strongly related, if it related above .40. Factors were weighted proportionally to number of items in the summing process. The resulting measures will be called paradigm factors, to distinguish them from the previous factors, although in a strict sense not all of them are factors.

When the paradigm factors were interrelated with each other, relationships between them across domains made it clear that although the paradigm was logically attractive, in some cases relationships across domains were quite strong. Where relations were as high as .70 or higher, and the factors seemed conceptually related, they were combined for study of process-product relationships. Three pairs of measures which related in the low .70's but were not clearly conceptually related were retained as individual factors, leaving 12 paradigm measures for further analysis. A measure of pupil interest in and attention to task was also included, even though it did not load on any of the factors, because of current interest.

Analysis of Outcome Measures within Grade

When the measures of pupil regressed gain (which had been created

separately within subgroups defined by sex, race, and socioeconomic status or IQ) were examined in preparation for reanalysis, we became aware that some first grade subgroups were several times more variable than others, for some measures. As a consequence, these data were transformed within subgroups to make them equally variable, so that some subgroups would not receive heavier weights than others in the final relationships. In addition, each subtest was examined for the possibility of ceiling effects or nonlinear relationships with pretests, and the Alphabet subtest was dropped from consideration as a gain measure because of serious ceiling effects.

Creation of a Pooled Data Set for First and Fifth Grades

In order to create a data set in which differences in classroom behavior were preserved from first grade to fifth grade when the process scores were created, area transformations were carried out item by item pooling data from first and fifth grade, so as to make the distributions for each item as normal and equally variable as possible, yet preserving differences from grade to grade. These transformed items were used to develop the paradigm factor scores described earlier.

The measures of pretest and regressed gain were transformed by pupil within grade level so as to make their distributions similar, and to set gain equal from grade level to grade level, since we were not confident that any differences in the data from different tests would present real differences in achievement from grade level to grade level. Similarly, pretest means were transformed within each grade for the same reasons.

One of the objectives was to compare classroom behaviors associated with gain for both high and low cognitive level outcomes. The subtests available were evaluated on a purely logical basis by thinking through the mental processes a pupil would use in answering the items in a subtest, and classifying them on the basis of the Taxonomy of Cognitive Objectives developed by Bloom et al (1956). A subtest was identified at each grade level which appeared to be at a high cognitive level, and another which appeared to be at a low cognitive level. Past analyses made that choice relatively easy for the fifth grade data--Vocabularly being the high cognitive level measure, and Spelling the low cognitive level measure. For first grade, the choice was not as clear, but Word Meaning was selected as the low cognitive level measure since all of the words used in the test were nouns, and required only recognition of the picture of an object, given its name ("Mark the bottle"). The Numbers subtest was selected as the high cognitive level measure since the items often required carrying out processes such as telling time, counting, or word problems, and using abstract terms such as "most", "larger", and "smaller". This was admittedly a subjective process, but on the other hand, the taxonomy was initially developed to classify test items in this way.

In addition to the high and low cognitive level measures, a sum of four achievement subtests was also obtained for each grade, as a coarse

measure of total achievement such as might be obtained from a test like the Wide Range Achievement Test. All four of the achievement tests from fifth grade were pooled; for first grade, in addition to Word Meaning and Numbers, Matching and Copying were used.

Process-Product Analyses

Few studies have taken account of pupil characteristics or setting variables in studying the relation of classroom behavior to pupil gain, and those which have only analyzed classroom means. One of the questions this study attempted to answer was whether taking such variables into account, within classroom, would identify classroom behaviors which were differentially effective. For example, within classrooms, do dependent pupils learn more than independent pupils if use of seatwork is high; and/or is the reverse true if seatwork is low?

Aggregating Gain by Subgroup. The way of examining this question which was proposed in the application was, within classroom, to aggregate the regressed gain of pupils at different levels of a characteristic such as dependency, so as to create subgroup means which would be more reliable than the individual pupil gain measure. That process was carried out for each gain measure for each of the pretest pupil characteristics by first using CROSSTABS from SPSS (Nie, Hull, Jenkins, Steinbrenner & Bent, 1970) to identify cutting points for all classrooms which would retain a minimum number of pupils within each subgroup within each classroom. Where possible, three subgroups were created, but in some cases only two subgroups could be created and still retain a minimum subgroup size. After the cutting points had been determined, then BREAKDOWN, also from SPSS, was used to calculate mean gain for each subgroup.

(Later work made clear that it was unnecessary to create these subgroups means first; that is if the cutting points were given SAS (Barr, Goodnight, Sall & Helwig, 1976) it would, in effect, aggregate the data internally in the analysis.)

Development of an Analysis for Relations Within Classroom. A central problem was the development of an analysis which would permit examination of relationships within classrooms. One of the unsettled questions in teacher effectiveness research is what the unit of analysis should be, but this is shared by other fields. Sociologists, political scientists and economists have typically analyzed aggregated data, have wished to generalize to individuals, but have been clear that such generalizations were hazardous or impossible (Dogan, M. & Rokkan, 1969). The recommended procedure from those fields has been to analyze data separately at each level of aggregation, but the problem of the nonindependence of observations from members of groups who interact and influence each other has not been confronted.

In much educational research all pupils who share a given treatment or innovation have been compared with pupils who experience other treatments, and the fact of membership in classrooms and schools has been ignored. In contrast, in research on teacher effectiveness, many researchers (Peckham, Glass & Hopkins, 1969, Glass & Stanley, 1970;

Wiley & Bock, 1967; Raths, 1967) have argued that the classroom is the only appropriate unit of analysis since pupils are not independent of each other within the same classroom. Peckham, et al. illustrate this clearly by the example of a pupil who is a discipline problem in a classroom, and who affects the learning process negatively for all other pupils in that classroom. But an analysis of classroom means assumes that the effect of a teacher behavior is the same for all pupils (or at least fails to recognize differences).

However, experienced teachers and some researchers argue that the same behavior may not be most effective for all pupils. This point of view is also basic to the Aptitude Treatment Interaction area of research (Cronbach and Snow, 1977), and the unit of analysis there has commonly been the individual, not the classroom group.

The basic issue is one of decomposing the total variation into portions attributable to between classroom means and within classrooms. It can be illustrated by considering two extreme possibilities for the placement of pupils within classrooms. As one possibility, assume that a school district assigns pupils to classrooms on a completely random basis (inconceivable, but proposed to make a point). In this case, within limits of sampling variability, all classroom means for pupil entering characteristics would be equal, and their interactions with teacher behavior would be zero. All relations between pre and post scores within classrooms would be lost to this analysis of classroom means.

As a second example, assume that the school system measures an entering characteristic such as IQ, and forms classrooms by assigning the highest standing 35 pupils to the first classroom, the next 35 to another classroom, and so on, so that an extreme example of system-wide homogeneous grouping has been created on this one pretest characteristic. In this case, variability between classroom means would be at a maximum for IQ, and an analysis using classroom means should identify a relationship between IQ and achievement and the interaction of IQ with teacher behavior if it exists (although, of course, other entering characteristics such as socio-economic status, anxiety level or motivation, which might be expected to influence response to teaching behavior, may be relatively unrelated to IQ and would still have considerable within-classroom variance).

The point to be made is that a part of the variance in pupil entering characteristics will ordinarily appear as between means (or between classrooms) variance and part as within classrooms variance. An analysis which deals only with classroom means misses some fraction of the important fraction of the total and could conceivably be the major source of variance. This is particularly likely where serious attempts at integration have mixed pupils from different neighborhoods.

The need for a satisfactory analysis to identify effects associated with both the between classrooms and the within classrooms variance in pupil entering characteristics seems compelling. A faculty member* of

*Dr. James T. McClave

the Statistics Department of the University of Florida (which is unusually strong in regression analysis) served on a continuing basis as a consultant in the development of such a procedure. This development was carried out over the summer of 1977, but when we did a "catch-up" review of the literature in the fall, we discovered that an abstract had appeared in RIE midsummer indexing a monograph by Cronbach (1976) in which he presented a solution to this problem.

We later realized that in the procedure which we had evolved we had reinvented the split-plot design, but that the parallelism had not been obvious since our analysis had involved the use of continuous measures across classrooms, using the general linear model, whereas the usual split-plot design has used classification variables and an analysis of variance computing procedure. However, it is not clear to us that Cronbach is employing the split-plot design in his recommended procedure, although the effect appears to be the same.

The proposed analysis is a two stage process in which variance in an outcome measure associated with measures of interest is first identified at the classroom level, followed by a similar analysis at the level of the pupil. The classroom level of analysis is completed by identifying variance for classrooms not identified by the previous measures.

The model identifies in successive steps the variance associated with a number of effects. For example, here are some questions which might be asked.

At the level of the classroom:

1) the linear effect of the pupil characteristic (e.g., does more learning occur in classrooms in which the average level of pupil motivation is high than in classrooms in which the average level of motivation is low?)

2) the nonlinear effect of the pupil characteristic (e.g., does more learning occur in classrooms with an intermediate level of average pupil motivation or is there a threshold or ceiling effect?)

3) the linear effect of the process measure (e.g., does more learning occur in classrooms where more time is spent in seatwork?)

4) the nonlinear effect of the process measure (e.g., does more learning occur in classrooms where an intermediate level of seatwork occurs, or is there a threshold or ceiling effect?)

5) the interaction of the pupil characteristic with the process measure (e.g., is the relation between motivation and learning different for classrooms where time spent in seatwork is great than for classrooms where it is little?)

6) the remaining classroom effect, which corresponds to the error term in a regression analysis of classroom means, or to variance for blocks in a split-plot design.

At the level of individual pupils, within classrooms:

7) the linear effect of the pupil characteristic (e.g., do highly motivated pupils learn more than low motivated pupils within classrooms?)

8) the nonlinear effect of the pupil characteristic (e.g., do pupils with an intermediate level of motivation learn more within classrooms or is there a threshold or ceiling effect?)

9) the interaction between class mean pupil characteristic and individual pupil characteristic (e.g., is the relationship between pupil motivation and learning different in classrooms where the class mean motivation is high than where it is low?)

10) the interaction of the pupil characteristic with the process measure (e.g., is the relation between pupil motivation and learning different in classrooms where time spent in seatwork is high than in classrooms where it is low?)

In summary, in the procedure which we evolved, the data should be analyzed both by classroom and by pupil, and we believe that the dependency problem is solved by using such a two stage analysis. In the first stage, relationships between classrooms are examined, paralleling the usual analysis of means in all respects except that the data of individual pupils are used as dependent variables rather than class means. The proportions of variance produced by this analysis are similar to those of the means analysis, but differ somewhat since each classroom receives a weight proportional to the number of pupils in it. Following the entry of all the variables of interest at the means level, all remaining variance associated with differences between classrooms is partialled out, creating the error term which is used to test between classroom effects (main plot error in the split-plot design). After that has been done, at the second stage, relationships are examined within classroom, using an error term made up of the remaining variance for individual pupils after all effects at the pupil level have been identified (sub plot error). If, then, membership in a particular classroom conveyed information about the average performance of pupils within that classroom, that information had been partialled out of the data before the second stage of the analysis.

As we understand it, the assumption of independence is not that observations themselves are independent, but rather that residuals are independent after the model has been fitted. Fitting a model to the data in which all classroom level variance has been partialled out then permits analysis of observations within classroom without violation of the assumption of independence. It is true that variance of pupils within classroom can be expected to be less than total variance of pupils, but this is a feature of the model which is parrallel to the reduction

of error variance which occurs in a repeated measures design.

A kind of random thought which has occured to us is that there is an intriguing irony in the parallel between the original meaning of the term "kindergarten", which means literally a garden in which children grow, and the realization that we were analyzing data within classroom using a split-plot design which was developed for agriculture.

Steps Followed in an Analysis. There were two steps in the analysis of each of the three data sets. The first was a screening process carried out by multiple regression; the second was a model building process carried out using GLM. For each of these, there was the two stage analysis of between class and within class effects.

The initial screening was carried out using stepwise regression to identify a smaller number of variables to be used in model building, since GLM will not select variables on the basis of variance accounted for. For the between class analyses, classroom means for pupil characteristics were used. For the within class analyses, the data of individual pupils were used, although this was rather imprecise since it represented total variance in common between process and outcome, rather than the within classrooms variance. As a consequence, a relatively larger number of variables was retained from the within class multiple regression. The variables subjected to screening will be reported in the Results chapter.

Following screening process, variables which were identified as strongest in terms of variance accounted for were used as the starting point for models to be tested using GLM (general linear model) from SAS (Barr et al., 1976), since it deals more efficiently with classification variables (i.e., the classrooms) than does multiple regression. In addition, it provides two estimates of the sums of squares accounted for by each variable; one is the sum of squares each variable accounts for as it enters the regression equation, which is order-dependent; the other is the sum of squares which each variable would have accounted for if it had been the last variable entered into the equation, which is not order-dependent. These latter sums of squares were the ones which were used in the selection of variables for reduced models and for interpretation of results.

Two approaches were taken in the model building process. In one, a model was developed to fit all outcome measures; in the other, a model or several models were developed which appeared to be "best fits" for each individual outcome measure. In those models, differences were sought in the teaching behaviors which related most strongly with pupil growth in outcome measures at different cognitive levels.

Chapter 4
Results and Discussion

Four general problems have been identified as the topics of study in this project: 1) to develop a set of process measures to serve as alternatives to the factor scores which have previously been used; 2) to develop a procedure for analysis of within classroom relationships; 3) to analyze interactions between characteristics of pupils and the teaching style which is most effective; and 4) to seek differences in the classroom behaviors which are associated with most effective learning for outcomes which differ in cognitive level. The results will be presented and discussed in the order of these topics.

Development of Process Measures

As was noted in the chapter on procedure, the use of individual items as measures of process was rejected because of the large number of relationships which would be tested and because of the usual low reliability of individual items. Factoring only items first found to relate to pupil gain was not pursued because of concern for representing only portions of constellations of teaching behavior. Instead, two procedures were followed: 1) identifying and scoring logically coherent subsets of items within the previously obtained factors; and 2) developing measures of classroom behavior based on a paradigm of the classroom environment for learning which grew out of rational analysis of previous empirical analyses of process-product relationships.

Logically Derived Subfactors

As indicated earlier, classroom behavior factors sometimes have presented interpretive problems. In an attempt to clarify the nature of the factors, item intercorrelations were examined for each of the eleven factors for the fifth grade sample, with the finding that for a number of the factors, clusters of items were present within each pole which were not highly interrelated. For example, one was Factor 1, Negative Control versus Orderly Classroom (Table 4.1). The pattern of items appears to be a reasonable one in which on the positive pole there are probably management problems in the classroom, and the teacher makes harsh attempts at control, both nonverbal and verbal, such as ordering and commanding, criticizing and scolding, glaring, frowning, holding or spanking, along with the use of negative affect. These behaviors contrast with the negative pole, which represents classrooms in which the emotional climate is positive and pupils are task-involved and self-directing. Although the factor related negatively to spelling gain, it was not clear what recommendation should be made to teachers, so the subfactors were scored and interrelated. The results are shown in Table 4.2. It is clear that relations are not consistently high between subfactor scores; subfactors within the positive pole related .35, within the negative pole, .17. Four of the six relations are lower than .40, which raises question about combining them into one measure, and leaves open the possibility of differential relations with pupil gain.

Table 4.1
Factor 1 - Negative Control vs Orderly Classroom[1]

Instrument	Subfactor[2]	Loading	Description
FLACCS	NB	.89	T negative affect, verbal and nonverbal
FLACCS	NB	.86	P negative affect, verbal and nonverbal
FLACCS	NB	.83	Strong T nonverbal control
FLACCS	NB	.80	Strong T verbal control
FLACCS	PF	.78	P engages in out-of-bounds behavior
RCS	NB	.78	T cools, formalizes
TPOR	NB	.70	T imposes external disciplinary control on P
FLACCS	PF	.61	P frequently socializes
RCS	NB	.57	T corrects
FLACCS	PF	.50	P aimless wandering
GR	OI	-.86	P self-control
FLACCS	OI	-.81	P interest-attention rating
FLACCS	OI	-.78	P follows routine without reminder
FLACCS	OI	-.77	P obeys directions
GR	EC	-.68	Overall emotional climate
GR		-.60	Cognitive focus
FLACCS	OI	-.53	P works, plays with little supervision
FLACCS	EC	-.49	T positive affect, verbal and nonverbal

Eigenvalue = 11.43

[1]After Soar & Soar (1973).

[2]NB = Negative Behavior EC = Emotional Climate
PF = Pupil Freedom OI = Order and Interest

Table 4.2
Subfactor Intercorrelations for Factor 1

	Negative Behavior	Pupil Freedom	Order, Interest
Pupil Freedom	.35		
Order, Interest	-.70	-.57	
Emotional Climate	-.31	.37	.16

Factor 2 (Table 4.3), Teacher Control, Varied Interaction was broken up into two subfactors, Teacher Central (TC) and Varied Interaction (VI), which correlated .69. The original factor was not related to gain, but since less than half of the variance in the two factors was in common, it seemed possible that one might relate to gain but be masked by the other.

Another example is Factor 5 (Table 4.4), Free Movement and Positive Affect with Little Focus. That factor had related negatively to reading and spelling gain, but it was not clear whether the lack of focus aspect (or the free movement) or the positive affect was the source of the relationship. When it was broken up into the two subfactors of Free Movement (FM) and Positive Affect (PA), they correlated .57, which would again leave considerable freedom for the subfactors to relate differently to gain.

Breaking the factors up into subfactors had the effect of approximately doubling the number of variables to be used in further analysis. This in itself is a problem of some seriousness, since the number of degrees of freedom in the study as a whole ranged from 49 to 59, depending on the outcome measure being analyzed. Since these measures might be expected to interact or to show nonlinear relations, the problem of degrees of freedom would be further aggravated.

The intercorrelations for the factors and subfactors are shown in Table 4.5. For Factor 1, it appeard that the strength of the factor lay primarily in Pupil Freedom (negatively), and secondarily, in Order and Interest. Factor 2 was neither improved nor changed by the breaking up into subfactors. For Factor 5, it was clear that the major strength of the factor lay in the measurement of movement (negative) rather than in affect. For both Factors 1 and 5 order and restraint of activity appeared to be more important than affect.

The logical analysis did appear to clarify relationships, but applied after factor analysis, it resulted in an objectionably large number of measures. It seemed potentially useful, then, to begin with a logical analysis, but one based on previous empirical analyses. The results of such an analysis follow.

Factor Based on a Paradigm for Learning

The other approach taken to the development of process measures consisted of identifying items from three of the observation instruments selected to represent aspects of the paradigm of the classroom environment for learning which was described on pp.12-14. The aspects for which factor analyses were carried out are as follows:

Table 4.3
Factor 2 - Teacher Central, Varied Interaction[1]

Instrument	Subfactor[2]	Loading	Description
RCS	VI	.79	T extended indirect
RCS	VI	.77	T-F flexibility
RCS	VI	.71	T accepts
RCS	VI	.69	T amplifies
RCS	VI	.65	P initiation following T indirect
TPOR	TC	.58	T center of attention
TPOR	TC	.48	T immediately reinforces P's answer as "right" or "wrong"
RCS	TC	.49	Drill
FLACCS	TC	.43	T central
RCS	TC	.44	T directs
Cog Tax		.41	T sum of interpretation
Eigenvalue = 6.52			

[1]After Soar & Soar (1973).
[2]VI = Varied Interaction TC = Teacher Central

Table 4.4
Factor 5 - Free Movement and Positive Affect with Little Focus[2]

Instrument	Subfactor[2]	Loading	Description
FLACCS	PA	.66	P positive affect,verbal and nonverbal
FLACCS	FM	.60	T moves freely among P
FLACCS	FM	.56	P seeks reassurance, support
FLACCS	FM	.55	T attends P briefly
FLACCS	PA	.52	T positive affect,verbal and nonverbal
FLACCS	FM	.48	P reports rule to another
FLACCS	FM	.42	P aimless wandering
FLACCS	FM	.43	P gives reason
FLACCS	PA	.42	Gentle T nonverbal control
Eigenvalue = 4.57			

[1]After Soar & Soar (1973).
[2]PA = Positive Affect FM = Free Movement

Table 4.5
Correlations of Factors and Subfactors with Pupil Gain
Fifth Grade[1]

Factor Number	Factor Title	Pupil Gain Vocabularly	Reading	Spelling	Arithmetic
1	Negative Control vs Orderly Climate	-.04	-.24	-.33*	-.10
	Negative Behavior	-.07	-.21	-.22	-.07
	Pupil Freedom	.09	-.31*	-.39**	-.05
	Order and Interest	-.00	.21	.40**	.10
	Emotional Climate	.06	-.06	-.01	.10
2	Teacher Central, Varied Interaction	-.24	-.00	.06	-.12
	Varied Interaction	-.23	-.03	.02	-.15
	Teacher Central	-.20	-.02	.10	-.03
5	Free Movement and Positive Affect With Little Focus	-.05	-.44**	-.30*	.04
	Positive Affect	.07	-.20	-.17	.09
	Free Movement	-.11	-.46**	-.32*	.00

[1] N = 49 classrooms

*P < .05

**P < .01

Teacher Emotional Climate
Pupil Emotional Climate
Teacher Management of Behavior
Pupil Management of Behavior
Teacher Management of Learning Tasks
Teacher and Pupil Management of Learning Tasks
Teacher Management of Thinking
Teacher Cognitive Level of Interaction

Following this, items from the Reciprocal Categories System were correlated with the factor scores from these factor analyses and added to the factors as the correlations indicated. Table 4.6 shows the RCS items which were added to the factors.

For the factor analyses of emotional climate data (Table 4.7), for both the teacher items and the pupil items negative affect and positive affect appeared in separate factors rather than being opposite poles of the same factor as emotional climate is often conceived. In both cases, both verbal and nonverbal behaviors appeared on each factor, and in both cases the eigenvalue for negative affect was greater than for positive affect, particularly for pupil negative affect. Related to this, when additional factors were rotated, the positive items were the ones which broke up, but the negative affect factors stayed together. Apparently negative affect as measured here was a strongly coherent constellation of behaviors.

For the Management of Behavior (Table 4.8) the teacher items broke up into one factor which represented primarily harsh, coercive direction-giving, and a second which was made up primarily of gentle, noncoercive behaviors, but with a single harsh behavior (scolds, punishes) with a low loading. Presumably, the first reflects a classroom with management problems; the second, a generally smooth running classroom with few management problems. The pupil behaviors also broke up into two factors. The first, Disorder versus Order, appeared to contrast much versus little socializing as an important aspect of the contrast between the positive and the negative poles, with restricted pupil activity on the pole representing order. The second factor, Follows Routine With Little Supervision, involves a variety of kinds of pupil activity which appear to be task oriented.

For the domain of Management of Learning Tasks (Table 4.9), the teacher behaviors created a factor in which the opposite poles contrasted classrooms in which the teacher is central, directing and monitoring the activities of the classroom, from those in which pupils are central and choose and carry out activities. There is also the distinction between pupils waiting, watching and listening, and pupils being actively involved. The second factor, Differentiated Activities, primarily represents a classroom where a variety of activities are going on simultaneously and pupils have some choice in what they will do and who they will do it with. Seatwork without the teacher is the strongest single item reflecting a task orientation in the traditional sense of

Table 4.6
RCS Items Added to Paradigm Factors

Paradigm Factor	RCS Item
Teacher Negative Affect	T cools, formalizes
Teacher Positive Affect	
Pupil Negative Affect	
Pupil Positive Affect	
Teacher Strong Control	T cools, formalizes
Teacher Gentle to Moderate Control	T warms, informalizes the climate
Pupil Disorder vs Order	Confusion
Pupils Follow Routine with Little Supervision	
Teacher Central and Directed	T elicits; T initiates; T directs; T corrects, P substantive interruption; Steady-state T talk
Differentiated Activities	P initiates
Interest Attention rating	
Teacher Guided Discovery Backed up by Facts	
Narrow, One-answer interaction	T narrow question; Drill
Pupils Guess or Hypothesize (no evaluation)	T broad question
T Low Level Cognitive Interaction	T talk
Teacher Mid to High Level Cognitive Interaction	T extended question

Table 4.7
Factor Analyses of Items for First and Fifth Grade Combined
Grouped by a Paradigm of the Classroom Environment for Learning

Emotional Climate

Instrument	Item	Loading
Teacher: Factor 1 Negative Affect (verbal and nonverbal)		
FLACCS	Says "stop it", etc.	.82
FLACCS	Frowns	.81
FLACCS	Uses threatening tone	.81
FLACCS	Shows disgust	.74
FLACCS	Points, shakes finger	.73
FLACCS	Yells	.61
FLACCS	Waits for child (negative)	.56
FLACCS	Rejects child	.54
FLACCS	Criticizes, blames	.53
	Eigenvalue 4.70	
Factor 2 Positive Affect (verbal and nonverbal)		
FLACCS	Smiles, laughs, nods	.71
FLACCS	Waits for child (positive)	.70
FLACCS	Listens carefully to child	.66
FLACCS	Touches, pats, hugs child	.65
FLACCS	Is enthusiastic, congenial	.61
FLACCS	Praises child	.56
FLACCS	Agrees with child	.56
FLACCS	Gives individual attention	.53
FLACCS	Develops "we" feeling	.49
FLACCS	Accepts favor	.44
	Eigenvalue 3.98	
Pupil: Factor 1 Negative Affect		
FLACCS	Teases	.83
FLACCS	Pushes or pulls, holds	.83
FLACCS	Threatens	.81
FLACCS	Picks at child	.79
FLACCS	Hits, hurts	.79
FLACCS	Makes disparaging remark	.76
FLACCS	Uncooperative, resistant	.75
FLACCS	Commands or demands	.74
FLACCS	Laughs (negative)	.70
FLACCS	Tattles	.69
FLACCS	Horseplay	.69
FLACCS	Interferes, threatens	.66
FLACCS	Makes face, frowns	.66
FLACCS	Finds fault	.61
FLACCS	Takes, damages property of others	.59

Table 4.7 (Continued)

Instrument	Item	Loading
Pupil: Factor 1 Negative Affect (cont'd)		
FLACCS	Says "no", etc.	.56
FLACCS	Makes someone "feel small"	.51
	Eigenvalue 8.80	
Factor 2 Positive Affect		
FLACCS	Sounds friendly	.72
FLACCS	Offers to share, cooperate	.72
FLACCS	Is enthusiastic (verbal)	.71
FLACCS	Helps another	.70
FLACCS	Is enthusiastic, happy(non-verbal)	.66
FLACCS	Helpful, shares	.64
FLACCS	Chooses another	.49
FLACCS	Frowns, pouts, withdraws	-.54
	Eigenvalue 3.81	

Table 4.8 (Continued)

Instrument	Item	Loading
Factor 2	Gentle to Moderate Control	
FLACCS	Praises	.77
FLACCS	Touches, pats	.76
FLACCS	Suggests, guides	.73
FLACCS	Directs with reason	.69
TPOR	Passes judgment on p's behavior or work	.64
FLACCS	Questions for reflective thought	.59
FLACCS	Uses time pressure	.52
FLACCS	Questions, states behavior rule	.49
TPOR	Discourages or prevents p. from expressing self freely	.46
FLACCS	Scolds, punishes	.41
	Eigenvalue 4.32	
Pupil: Factor 1	Disorder vs. Order	
FLACCS	Frequently socializes	.94
FLACCS	Aimless wandering	.78
FLACCS	Engages in out-of-bounds behavior	.75
FLACCS	Gives direction	.50
FLACCS	Reports rule to another	.45
FLACCS	Occasionally socializes	-.69
FLACCS	Almost never socializes	-.68
FLACCS	Obeys directions	-.60
	Eigenvalue 3.96	
Factor 2	Follows Routine with Little Supervision	
FLACCS	Follows routine without reminder	.78
FLACCS	Gives reason	.74
FLACCS	Works, plays with little supervision	.52
FLACCS	Obeys directions	.42
FLACCS	Reports rule to another	.40
	Eigenvalue 1.96	

Table 4.8
Factor Analyses of Items for First and Fifth Grade Combined Grouped by a Paradigm of the Classroom Environment for Learning Management of Behavior

Instrument	Item	Loading
Teacher: Factor 1	Strong Control	
FLACCS	Gives orders, commands	.77
FLACCS	Glares, frowns	.73
FLACCS	Uses sharp tone	.72
TPOR	Imposes external disciplinary control on p.	.72
FLACCS	Scolds, punishes	.64
FLACCS	Interrupts pupil, cuts off	.62
FLACCS	Directs without reason	.61
FLACCS	Holds, pushes	.60
TPOR	Discourages or prevents p from expressing self freely	.59
TPOR	Passes judgment on p's behavior or work	.51
TPOR	Withholds judgment on p's behavior or work	-.59
TPOR	Encourages p to express self freely	-.55
TPOR	Encourages self discipline on part of p.	-.54
	Eigenvalue 5.45	

Table 4.9
Factor Analyses of Items for First and Fifth Grade Combined Grouped by a Paradigm of the Classroom Environment for Learning Management of Learning Tasks

Instrument	Item	Loading
T & P: Factor 1	Teacher Central and Directed vs. Pupil Central and Active	
TPOR	Occupies center of attention	.75
TPOR	Provides p. with detailed facts & info.	.67
FLACCS	Central	.65
FLACCS	Group with Teacher	.65
TPOR	Approaches subject matter in direct, business-like way	.61
TPOR	Makes some thing as a thing center of p's attention	.60
TPOR	Has p spend time waiting, watching, listening	.57
FLACCS	Ignores pupil, refuses to attend	.42
TPOR	Organizes learning around p's own prob.	-.64
TPOR	Has p make his own collection & anal. of subject matter	-.62
FLACCS	Pupil central	-.45
FLACCS	Pupil task related movement	-.44
TPOR	Makes doing something center of P's atten.	-.42
	Eigenvalue 5.39	
Factor 2	Differentiated Activities	
FLACCS	Seatwork without the teacher	.70
TPOR	T has diff. P working at diff. tasks	.67
FLACCS	T attends simultaneous activities	.66
FLACCS	P speaks aloud without permission	.63
TPOR	T has P work indepen. on what concerns P	.59
FLACCS	P works, plays collaboratively	.58
TPOR	T joins or participates in P's activ.	.56
FLACCS	P--limited choice	.53
FLACCS	Free groups	.52
FLACCS	T attends P closely	.51
FLACCS	P--free choice	.50
TPOR	T has p participate actively	.48
TPOR	T makes a wide range of info. mat. avail.	.41
TPOR	T relies heavily on testbook as source of info.	-.48
FLACCS	Seatwork with teacher	-.47
TPOR	T holds all P responsible for certain material to be learned	-.41
FLA		
	Eigenvalue 5.35	

the word. Whereas the previous factor contrasted who was central, teacher or pupil, no one is front and center in the second factor; rather, the teacher moves from pupil to pupil through several activities.

For the Management of Thinking (Table 4.10), two factor analyses were carried out, one which involved items reflecting the degree to whcih the teacher directs and manages the thinking processes, and another which reflected the cognitive level at which the thinking processes were to be carried out. During the rational analysis in which item pools were being established, it was not entirely clear whether this first set of items belonged in Learning Tasks or Thinking, since the thinking process is embedded in a learning task. The items seem to reflect thinking to a greater degree than learning tasks, and consequently they were classrified in this domain, but they are to a degree transitional.

The items reflecting teacher-management of the Thinking Processes broke up into three factors. In the first, Guided Discovery Backed up by Facts, the teacher encourages pupils to go beyond "the facts given" but prominent in the factor are items which reflect testing and evaluating the new ideas, in which the pupil is actively involved. The second factor, Narrow, One-Answer Interaction, reflects a classroom in which teacher-pupil interaction is narrowly restricted to pupils providing the one right answer which the teacher has in mind and which comes directly from the lesson. The third factor, Guess or Hypothesize (No Evaluation), like the first, encourages pupils to suggest new and different ideas, but unlike the first, does not engage pupils in testing or evaluating, or backing up the ideas with facts. In our work with the factor, we have characterized it as "loose and sloppy" in the sense that any answer is all right and there is no hard thinking. Perhaps "brainstorming" would be a better term, an activity which has its value, but the thinking process must go beyond this to be rooted in reality.

The items reflecting cognitive level of interaction included only teacher items since pupil items were so highly related as to be redundant. They broke up into two factors. The first was Low Cognitive Level Interation, in which the two strong loadings were the two lowest levels of the taxonomy, reflecting memory level activity and changing the form in which an idea is expressed without changing the idea or information itself. Synthesis, Level 6, also appears, but with a minimal loading. The second factor, Mid to High Cognitive Level Interaction involved Levels 3, 4, and 5 of the taxonomy, which reflect pupils carrying out previously learned processes, deciding what procedures will solve a problem and carrying them out, and inferring and self-conscious logical processes.

The interrelationships between the paradigm factors, revised to include selected RCS items, are high in a scattering of instances (Table 4.11), but are generally low to moderate. The high correlations did suggest combining measures where they were conceptually related, so this was done for Teacher Negative Affect and Strong Control, which correlated .91; for Pupil Negative Affect and Disorder, which correlated

Table 4.10

Factor Analyses of Items for First and Fifth Grade Combined Grouped by a Paradigm of the Classroom Environment for Learning Management Thinking

Instrument	Item	Loading
Teacher: Factor 1 Guided Discovery Backed up by Facts		
TPOR	Involves P in uncertain or imcomplete situation	.73
TPOR	Asks P to judge comp. value of answers or ques.	.62
TPOR	Asks P to support answers or opinion with evidence	.61
TPOR	Helps P discover and correct factual errors and inaccracies	.59
TPOR	Encourages P to put his ideas to test	.58
TPOR	Asks P to eval. his own work	.56
TPOR	Leads P to Q or prob. which "stumps" him	.53
TPOR	Quest. misconceptions, faulty logic, unwarranted conclusions	.52
TPOR	Has P decided when Q has been answered satis.	.49
TPOR	Motivates P with intrinsic values of ideas or activities	.44
TPOR	Permits P to suggest add. or alter. answers	.43
	Eigenvalue 4.05	
Factor 2 Narrow, One-Answer Interaction		
TPOR	Expects P to come up with answ. T has in mind	.84
TPOR	Immed. reinforces P's answ. as "right" or "wrong"	.77
TPOR	Accepts only one answer as being correct	.70
TPOR	Prevents sit. which cause P doubt or perplexity	.65
TPOR	Accepts only answers or suggest. closely related to topic	.62
TPOR	Evaluates work of all by a set standard	.58
TPOR	Stops P from going ahead with plan which T knows will fail	.55
TPOR	Asks Q that P can answ. only if he studied the lesson	.54
TPOR	Expects P to "know" rather than quess answer to Q	.52
TPOR	Asks another P to give answ. if one P fails to answer quickly	.52
TPOR	Emphasized idealized, reassuring, or "pretty" aspects of topic	.46
	Eigenvalue 5.19	

Table 4.10 (Continued)

Instrument	Item	Loading
	Factor 3 Guess or Hypothesize (no evaluation)	
TPOR	Asks Q that is not readily answerable by study of lesson	.72
TPOR	Permits P to suggest additional or alternative answers	.69
TPOR	Encourages P to guess or hypothesize about unknown or untested	.59
TPOR	Emphasizes idealized, reassuring, or "pretty" aspects of a topic	.56
TPOR	Entertains even "wild" or farfetched sugg.	.53
TPOR	Emphasized realistic, disconcerting, or "ugly" aspects of topic	.48
TPOR	Steers P away from hard Q or problem	.41
TPOR	Motivates with intrinsic value of ideas or activity	.42
TPOR	Expects P to "know" rather than to guess answer to Q	-.64
TPOR	Asks Q that P can answer only if he studied the lesson	-.46
TPOR	Accepts only one answer as being correct	-.45
	Eigenvalue 3.96	
Teacher: Factor 1	Low Cognitive Level Interaction	
Cog Tax	Sum of Translation	.85
Cog Tax	Sum of Knowledge	.84
Cog Tax	Synthesis	.40
	Eigenvalue 1.71	
	Factor 2 Mid to High Cognitive Level Interaction	
Cog Tax	Sum of Interpretation	.76
Cog Tax	Sum of Analysis	.72
Cog Tax	Sum of Application	.71
	Eigenvalue 1.68	

Table 4.11
Correlations Between Measures Based on a
Paradigm of the Classroom Atmosphere for Learning Grades 5 and 1

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	1
Emot Climate:																				
T Neg	1	--																		
T Pos	2	-35	--																	
P Neg	3	51	07	--																
P Pos	4	-33	70	21	--															
Behavior:																				
T Strong Control	5	91	-42	47	-38	--														
T G to Mod Control	6	44	39	47	20	43	--													
P Disorder vs Order	7	34	02	77	37	25	16	--												
P Follow Routine	8	-46	60	-03	64	-50	22	-03	--											
Learning Task:																				
T Central & Directed	9	47	-12	19	-22	60	48	-18	-17	--										
Diff. Activities	10	-12	56	31	72	-19	40	36	50	-21	--									
Thinking:																				
T Guided Disc	11	-24	57	-10	47	-29	36	-06	44	-02	40	--								
T Narrow, One Ans	12	32	-09	-02	-20	47	40	-28	-03	70	-18	14	--							
T Enc Guess, Hyp	13	-02	53	29	40	-14	53	14	38	07	56	54	-17	--						
Low Level Cog	14	26	25	18	10	30	63	-15	23	72	21	24	57	40	--					
Mid to High Cog	15	03	-43	-50	-44	05	-37	-37	-40	14	-51	-04	23	-34	-06	--				
Pooled:																				
T Neg & Strong Control (1+5)	16	97	-40	51	-36	98	44	30	-50	56	-16	-28	41	-08	28	04	--			
P Neg & Disorder (3+7)	17	48	06	97	28	42	40	89	-03	07	34	-08	-11	26	08	-48	47	--		
Interest-Attention	18	-62	51	-42	51	-65	-07	-34	57	-30	39	40	-19	27	-01	-04	-65	-42	--	
T & P Positive Affect (2+4)	19	-36	.91	.17	.93	-.42	.30	.24	.66	-.14	.69	.57	-.13	.50	.21	-.48	-.40	.21	.55	--

.77; and Teacher and Pupil Positive Affect, which correlated .70. This left 12 paradigm factors (revised), plus the measure of pupil Interest and Attention to task, which was carried forward as interesting in its own right. Several other factors which were correlated above .70 were not combined because conceptually they represented somewhat different measures. There were Differentiated Activities and Pupil Positive Affect, which related .72; Teacher Narrow One-answer Interaction and Teacher Central and Directed, which related .70; and Low Level Cognitive Interaction and Teacher Central and Directed, which were correlated .72. The other correlations seemed low enough as not to suggest further combining.

The paradigm factors seemed to us to be preferable to the subfactors since the paradigm factors were less numerous and since they were more conceptually distinct. In contrast, there were subfactors which seemed to be similar conceptually, as for example, Pupil Free and Free Movement which were subfactors of Factors 1 and 5.

In addition, preliminary multiple regressions which compared the paradigm factors with the original eleven factors showed that although the factors accounted for somewhat more variance in the outcome measures between classes, the paradigm factors appeared to offer clearer interpretation. Accordingly, the paradigm factors were used in analyzing process-product relationships. The intercorrelations of the paradigm factors with gain are shown in Appendix A.

Development of the Method of Analysis

Since the analysis of within classroom effects had to be carried out using either individual pupils or subgroups of pupils, an initial question was whether, if individual pupils were used in the between classroom analyses, the results would parallel those of the usual analysis of means. Table 4.12 presents a comparison of the same variables analyzed using means in the usual way and also using outcome measures for individual pupils. It can be seen that the results are quite similar for the two sets of data which are presented. It seems clear that when a process measure or a pretest mean has the same value for the independent variable for all pupils in a classroom, the effect is similar whether pupils or means are used for the dependent variable. The differences that do appear presumably occur because the analysis using individual pupils weights each classroom proportionally to the number of pupils, whereas the means analysis gives each classroom equal weight regardless of the number of pupils. Not only does this proportional weighting seem preferable in itself, it also eliminates the problem of setting some minimum number of pupils to include a classroom in the analysis. The proposal on which this research is based had suggested aggregating the data of subgroups of pupils within each classroom so as to create subgroup means which would be more reliable than individual scores, on the assumption that this procedure would result in greater proportions of variance accounted for. The comparison of means and individual pupils suggests that this is not so.

Table 4.12

Comparison of Classroom-Level Effects for Vocabulary Gain Calculated from Classroom Means and from Individual Pupils

Fifth Grade*

Source	Means SS	Means %V	Pupils SS	Pupils %V
Seatwork (A)	29.47	9	599.09	8
Seatwork^2	.47	0	14.25	0
Motivation (B)	20.38	6	555.95	7
Motivation^2	21.98	7	431.10	8
A x B	16.67	5	289.68	4
Error	343.92	73	7475.63	73
Recitation (A)	.42	0	1.58	0
Recitation^2	91.61	32	2104.12	34
Motivation (B)	17.51	6	420.41	7
Motivation^2	37.82	13	667.22	11
A x B	.07	0	5.78	0
Error	285.47	49	6166.60	48

*N = 59 Classrooms

Table 4.13 reports results for analyses in which means for subgroups were used, in which the analysis procedure was given cutting points to aggregate subgroup data internally, and in which the analysis was carried out with individuals. It can be seen that the proportions of variance accounted for by the analysis of subgroup means are quite different from those of the other two analyses, which are highly similar to each other. The subgroup means are in a sense only one step removed from classroom level variance, and the total sum of squares is relatively small. But the other two analyses use the pupil as the unit of analysis so that the total sum of squares is much larger. The different bases, then, create percentages of quite different orders. In addition, subgroup means are equally weighted, regardless of the number of pupils, so that idiosyncratic results seem likely. Since aggregating pupils within the analysis, which weights each subgroup proportionately to the number of pupils in it, produces results highly similar to those of the analysis of individuals not aggregated, it appears to be simpler and just as useful to analyze individuals. Accordingly, this was the procedure which was followed in later analyses.

Process-Product Analyses for First Grade

Screening

As noted in the Procedure Chapter, first steps in the process-product analyses were two stepwise multiple regressions, one for Between Classrooms, and one for Within Classrooms. Their purpose was to give a rough indication of the variables which had most variance in common with an outcome measure so that a smaller number of variables could then be tested using SAS GLM. Again, GLM will not select variables in the order of variance accounted for, but multiple regression will not easily deal with classrooms as classification variables, so the analyses were used in this order to take advantage of the strengths of each.

Table 4.14 reports the variables which were screened for the Between Classroom analyses. Twenty variables were screened, which are too many to permit drawing dependable conclusions, and as a consequence, the results can only be depended upon as they are replicated in other studies. However, in an area where little is known, where theory is as often misleading as helpful, and where data collection is very expensive, the concern for failing to examine relationships of possible interest seemed to us to be compelling. Even with the number of variables which were included, numbers of others of potential interest were not studied.

In interpreting the table, the column headed "Sq" indicated squared terms which tested nonlinear relations. The column headed "Pdm" lists the number of the paradigm process measure which was interacted with the paradigm measure indicated by the row; other columns under "Interaction With" indicate interactions between pupil characteristics (in this case means) and classroom process.

Table 4.13

Comparison of Three Ways of Analyzing
Within Class Relationships

First Grade*

Source	Previously Calculated Subgroup Means		Individuals Aggregated Internally		Individuals not Aggregated	
	SS	%V	SS	%V	SS	%V
Hard Question (A)	43.39	6	534.81	1	534.81	1
Hard Question^2	106.72	16	1097.40	2	1097.40	2
Classrooms	295.63	44	2716.89	6	2716.89	6
Pupil SocioEconomic Status (B)	.03	0	103.41	0	.20	0
A x B	.79	0	.68	0	27.14	0
Error	225.82	34	44569.09	91	44645.84	91

*N = 22 classrooms, 503 pupils

Table 4.14

Variables Screened, Between Classrooms
First Grade[1]

Measure	First Order	Interaction With			
		Sq	Pdm	SES	PreT	Dr-a-M
Paradigm Factor						
P6 T Gentle to Moderate Control	X			X		
P11 Teacher Central	X					
P12 Differentiated Activities	X	X				X
P15 Interest-Attention Rating	X				X	
P17 Narrow, One-Answer	X				X	
P18 Pupils Guess	X					
Pooled						
P19 T Strong Control & Neg.	X			X		
P20 P. Negative & Disorder	X					
P21 T & P Positive Affect	X					
Pupil Characteristics--Class mean						
Pretest	X					
Draw-a-Man	X					
Familiar Figures	X					
SES	X					
Race	X					

[1]N = 22 classrooms

The variables which were screened for Within Classroom analysis are reported in Table 4.15. Entries are not shown for the paradigm measures as first order, squared, or interaction terms with other process measures, since all of these are Between Class effects. The entries in this table do reflect first order relationships between pupil characteristics and gain within classroom, nonlinear relations for pupil characteristics, and interactions between individual pupil characteristics and classroom process. Although 39 variables were entered into this analysis, the problem of the number of variables being too large for the degrees of freedom does not exist here since over 500 degrees of freedom were available.

From these screening analyses, variables were selected for building models for predicting each of the outcome measures, both between and within classrooms, using a rather liberal cut-off which carried forward a relatively large number of variables for further testing. Although results at the Between Class level were not the primary focus of this project, they are reported since this new analysis weights each classroom by the number of pupils in it, since it seemed possible that the new paradigm factors would result in more interpretable process-product relationships and most importantly since it seemed useful to compare relative amounts of variance accounted for between classes and within classes.

The results of the screening analyses were used to identify variables which seemed most promising for inclusion in models to be tested at the two levels of Between and Within Classrooms using SAS GLM. For each outcome measure, the attempt was made to identify a model made up only of first order terms and, where it seemed appropriate, also one which included second order terms. In some cases, more than one model of these kinds appeared to be worth testing in the attempt to identify as much variance as possible with a small number of variables.

Models

Word Meaning. The results of testing models for Word Meaning are shown in Table 4.16. In that table and others which follow, the numbers in the body of the table are percentages of variance in common between independent and dependent variables, and signs for first order variables are interpreted as they are for correlations. For example, in the 4 variable model in Table 4.16, Teacher Central accounted for 17% of the variance in gain in Word Meaning, and the relation was negative, significant at the five percent level. For second order variables representing interactions, the sign indicates the direction of the change in slope for the regression of one independent variable (IV) with the outcome as the other IV goes from low to high. For the example below, the sign

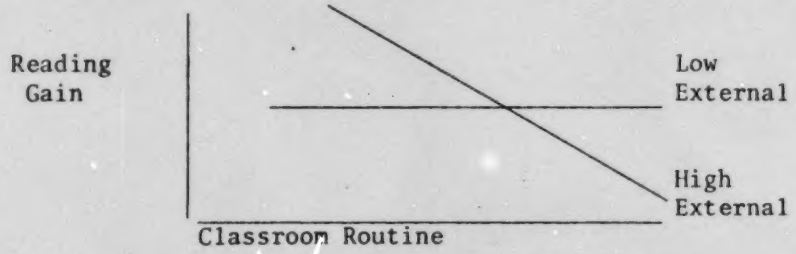


Table 4.15

Variables Screened, Within Classroom
First Grade[1]

Measure	First Order	Sq	Interaction With: Pre	Alph	Dr-a-m	FF	SES	Sex	Race
Paradigm Factor									
6 T Gentle to Moderate Control							X	X	
8 Pupils Follow Routine			X	X	X	X	X	X	X
12 Differentiated Activities			X	X	X	X	X		
15 Interest-Attention Rating			X	X	X	X	X		
17 Narrow, One-Answer									X
18 Pupils Guess			X	X	X	X			
9 Low Cognitive Level							X		
Pooled									
19 T Strong Control & Neg. Aff.								X	X
20 P. Negative & Disorder							X	X	X
21 T & P Positive Affect							X		X
Pupil Characteristics									
Pretest	X								
Alphabet	X								
Draw-a-Man	X								
Familiar Figures	X								
SES	X								
Sex	X								
Race	X								

[1]N = 503 pupils

Table 4.16

Variance in Word Meaning Accounted for,
First Grade

Measures	Models	
Between Classrooms[1]	4 Variable	3 Variable
P11 T Central	17-*	16-*
P18 P Guess	04+	
P21 T & P Pos. Affect	29-**	28-**
Draw-a-Man	43+**	39+**
Model F	7.14**	8.50**
Within Classrooms[2]	3 Variable[3]	1 Variable[3]
Pre Alphabet	.8+*	1.3+*
Familiar Figures	1.1+*	
Differ. Activ. x Fam. Fig.	.9-*	
Model F	4.38**	6.49*

*P < .05 **P < .01

(1.1+*) First Order term not interpretable since the variable is a member of a second order term

[1]N = 22 classrooms

[2]N = 503 pupils

[3]For comparability with Between Class variance, multiply each percent by 11.2.

would be negative indicating that the relation between classroom routine and gain becomes more negative (or less positive) as externality goes from low to high. A positive sign would indicate a change in slope becoming more positive (or less negative) as the other IV goes from low to high. The information about direction and degree of slope of the two regression lines is sacrificed for the sake of compactness of presentation. Model F is the significance test for the model as a whole.

In each case the percent of variance reported is the amount which would be credited to the variable if it had been the last to enter the equation, and as a consequence the results are not order-dependent.

The percentages of variance reported for the Between Class analyses use classroom level variance as the base and therefore are comparable to the usual analyses of class means. However, the total variance accounted for by differences between classrooms typically ranges between about eight and 15 percent of the total sum of squares, which means that the Within Class variances which are based on the total sum of squares are on a quite different scale. To be comparable, they would need to be multiplied by constants ranging from six to 13, depending on the outcome measure.

For the Between Classrooms models for Word Meaning, the three variable model appears to be preferable in that nearly as much variance is accounted for but with one fewer variables. The most powerful predictor is class mean Draw-a-Man, a rough measure of intelligence, followed by expression of positive affect in the classroom, which was negatively related to gain. While the finding for positive affect is contrary to widely held beliefs, it may be because the items in the measure do not reflect task involvement, and perhaps also because more affect may be expressed in classrooms in which there is greater freedom of behavior, which in turn tends to be negatively related to achievement. Results for positive affect are mixed in the Medley (1977) review, and praise, as one aspect of positive affect, has been shown to relate negatively in some circumstances by Rowe (1974) and Brophy and Evertson (1974).

The negative relationship for Teacher Central indicates that less learning occurred in classrooms where the teacher often took a central role in assigning and directing learning activities. This appears to be a result which differs from the recommendations for direct instruction (Rosenshine and Berliner, 1977).

Within Classrooms, in a single variable model, pretest Alphabet was significantly related with gain. Although it accounted for about one percent variance, the percent comparable to the Between Class results would be 15. This measure was chosen for inclusion as a predictor on the expectation that a child who scored high in naming the letters of the alphabet was more likely to have support for learning at home than a child who scored low. In the three variable model, Alphabet predicted almost as strongly as in the single variable model, but Familiar figures and

its interaction with Differentiated Activities accounted for another two percent of variance (comparable to 22 percent). The interaction indicated that there was greater gain for reflective pupils than for impulsive pupils in classrooms where differentiation was low. Perhaps the higher activity level of the differentiated classroom interferes with the preferred learning process of the reflective child (Fig. 4.1).

On the surface, the most striking aspect of this table is the greater variance accounted for Between Class as compared to Within Class. The different bases on which the percentages were calculated is the major difference, but the Between Class percents maybe inflated to a degree by the large number of measures brought to the analysis. What does seem clear is that the Within Class analysis produces a meaningful amount of variance, whether due to pupil characteristics or interaction with process.

Listening. For Listening, (Table 4.17) among the models tested at the Between Classrooms level, the choice appears to be between the two variable model and the five variable model since the variable added for the three variable model was not significant. In all three models classroom behavior measures are the major source of variance accounted for, with Teacher Strong Control and Negative Affect relating strongly negatively with gain, and Differentiated Learning Activities relating negatively but less strongly. In this sense, the two variable model makes a clear statement about less effective teaching. The five variable model adds a significant positive relationship between Familiar Figures and gain, indicating that greater gain was shown in classrooms where pupils were more reflective, on the average. Since that relationship was not significant in the three variable model but became so in the five variable model, it appears to have been strengthened when the effects of mean pretest and mean SES were partialled out.

Suppressor effects are frequent in these data and this appears to be an example. In the multiple regression literature, the problem of colinearity is frequently mentioned--the problem of two or more independent measures sharing variance in common with the dependent measure, so that as variables are added to the regression equation the variance accounted for by each is reduced. In contrast, when the variance in common between an independent and the dependent variable increases as variables are added to the regression equation, suppression effects are suspected. Cohen and Cohen (1975) deal with this problem more extensively than most sources, identifying the simplest case as one in which the second independent variable partials out or suppresses variance in the first independent variable which is unrelated to the dependent variable.

For the Within Classroom results the stronger predictors are pupil characteristics. The two variable model again is relatively strong, indicating that pupils who came to school knowing the alphabet gained more during the year than those who came with low scores. The negative relation for pretest is presumably a reflection of pretest having been partialled out at the level of the individual pupil when regressed gain scores were calculated, and that classroom level effects have been partialled out before this result. The five variable model does add an interaction with classroom process, however, which indicates that where Teacher Strong Control and Negative Affect were high, white pupils

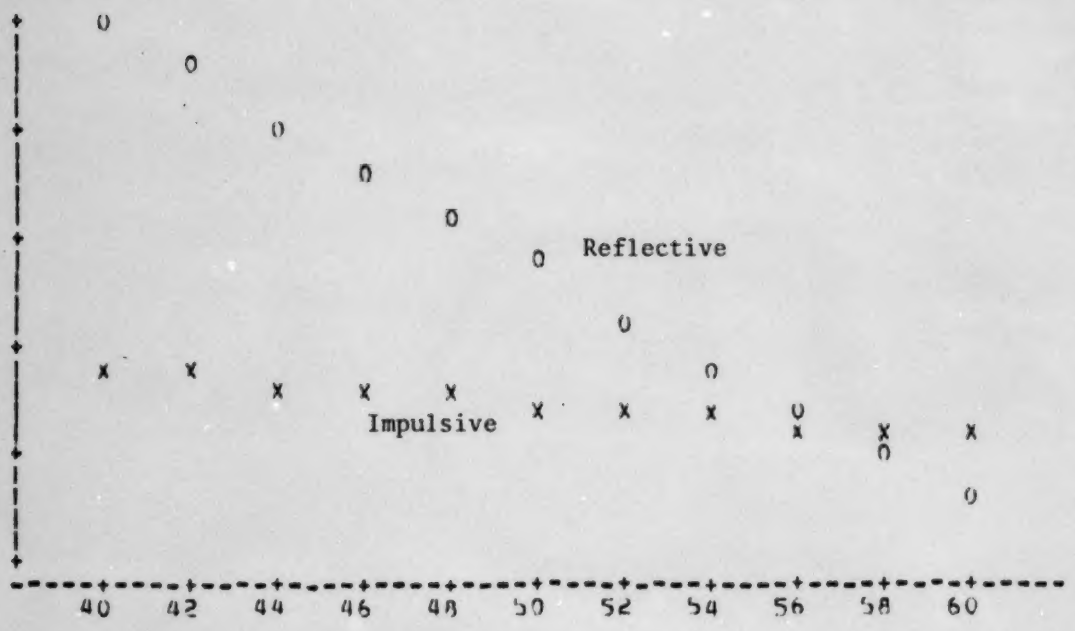


Figure 4.1 Interaction of Differentiated Activities with Familiar Figures for Word Meaning Gain, Within Class, for First Grade

Table 4.17

Variance in Listening Accounted for,
First Grade

Measures	Models		
Between Classrooms[1]	5 Variable	3 Variable	2 Variable
P12 Differ. Activ.	14-*	14-*	16-*
P19 T Strong, Negative	50-**	45-**	51-**
Mean Pretest	4-		
Mean Familiar Fig.	15+*	7+	
Mean SES (1-high, 7-low)	9+		
Model F	5.78**	8.30**	10.01**
Within Classrooms[2]	5 Variable[3]	2 Variable[3]	
Pretest	1.8-**	2.2-**	
Alphabet	4.5+**	4.7+**	
Familiar Figures	.5+		
Race (1=Black, 2=White)	1.5-**		
Race x T Strong, Neg.	1.2+*		
Model F	7.47**	12.73**	

*P < .05 **P < .01

[1]N = 22 classrooms

[2]N = 503 pupils

[3]For comparability with Between Class variance, multiply each percent by 8.7.

◯ First Order term not interpretable since the variable is a member of a second order term

learned more than black pupils; where there was little strong control, black pupils learned more than white. At the risk of oversimplification, black pupils seemed to be more hindered by teacher strong control than white pupils(Fig.4.2).

Matching. For the Between Classes analysis for Matching (Table 4.18), only one variable was related, Gentle to Moderate Control. The relationship indicates that gain was less in classrooms where the teacher used more classroom management statements that ranged from gentle and unobtrusive to routine directions. Perhaps where this style of management of behavior is frequent, less teaching is occurring.

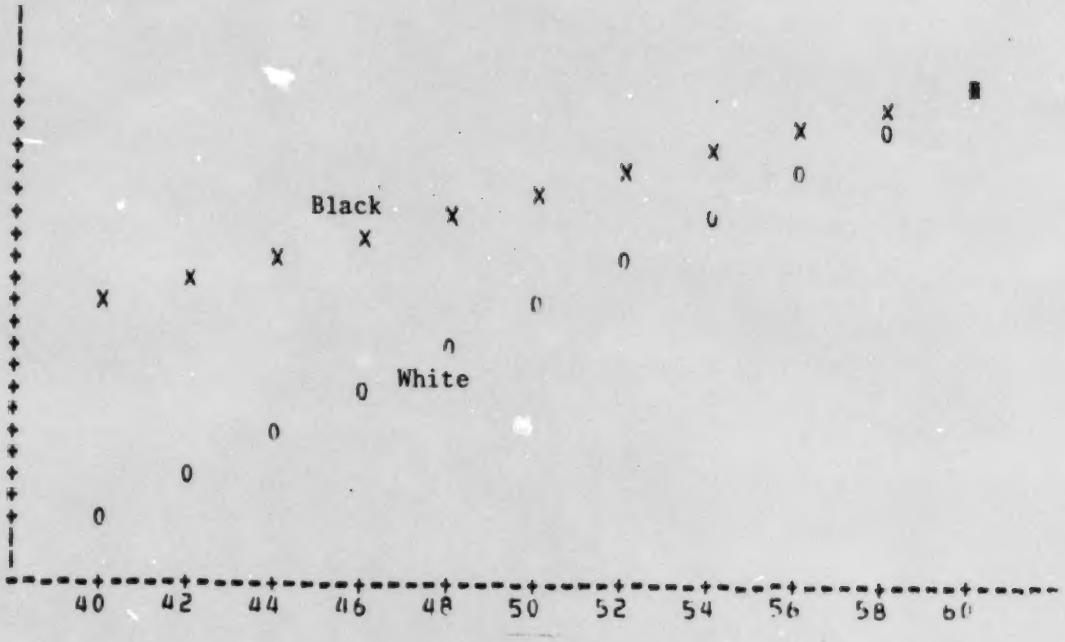
For the Within Class results, it appears that the three variable model is preferable. The six variable model adds a small increase in variance accounted for, but adds three interactions to bring it about. The briefer model includes pretest, negatively related, presumably because of the previous partialling of classroom level effects. In addition to it, both Alphabet and Familiar Figures contribute large amounts to variance which is accounted for at the Within Class level.

Numbers. The Between Classrooms analysis for Numbers (Table 4.19) tested three models, of which the four variable appears to be preferable. Although it adds several variables, it is made up entirely of first order terms, is the only model which is significant, and accounts for considerably more variance than either of the other two. The strongest variables in that model, and the only ones individually significantly related, are mean Draw-a-Man, the rough intelligence measure, which was positively realted to gain, and mean SES. As indicated in the table, the scale for SES was reversed, so that the sign indicates greater gain for low SES pupils. SES was originally adjusted out of the gain measures, permitting all the other variables in this model to vary. Although frankly speculative, perhaps, then, partialling out the relation of Draw-a-Man to gain, which was strongly positive, had the effect of showing greater gain for low SES pupils.

Although the relationship for Interest-Attention was not quite significant, the fact of the negative relationship may be worth note since this outcome measure was earlier judged to be most complex.

For the Within Class analysis, the model using only first order terms appears to be preferable. Not only did it avoid more complex relationships, but every variable in it was significant. Pretest showed the same negative relation it had shown for other outcome measures, presumably for reasons cited earlier. In addition, Alphabet, Draw-a-Man and Familiar Figures all related positively and significantly with gain, with considerable amounts of variance accounted for. As in the previous analyses, the effects of classroom behavior appeared for the Between Classes analyses rather than as interactions in the Within Classes analyses, but variance accounted for within classrooms was comparable to that between classrooms.

Copying. In the Between Classes analysis for Copying (Table 4.20),



Teacher Strong Control & Negative Affect

Figure 4.2 Interaction of Teacher Strong Control and Negative Affect with Race for Listening Gain, Within Class, for First Grade

Table 4.18

Variance in Matching Accounted for,
First Grade

Measure	Model	
Between Classes[1]	1 Variable	
P6 Gentle to Mod. Cont.	24-*	
Model F	6.34*	
Within Classes[2]	6 Variable[3]	3 Variable[3]
Pretest	3.1-**	2.8-**
Alphabet	(1.6+**)	1.8+**
Fam. Figures	(0.0+)	2.0+**
F F x Differ. Activ.	.7-	
Alphabet x P. Routine	.9-*	
F F x Interest-Attn.	.9+*	
Model F	5.58**	8.03**

*P < .05 **P < .01

( ) First Order term not interpretable since the variable is a member of a second order term

[1]N = 22 classrooms

[2]N = 503 pupils

[3]For comparability with Between Class variance, multiply each percent by 10.4.

Table 4.19

Variance in Numbers Accounted for,
First Grade

Measures	Models		
Between Classrooms[1]	4 Variable	3 Variable	1 Variable
P12 Differ. Activ.	4-	(4+)	
P15 Interest-Attention	10-		
Mean Draw-a-Man	23+*	(3+)	7+
Mean SES (1-High; 7-Low)	16+*		
Diff. Activ. x Dr-a-M		7-	
Model F	3.01*	2.30	1.40
Within Classrooms[2]	First Order[3]	Second Order[3]	
Alphabet	2.8+**		
Pretest	3.8-**		
Draw-a-Man	.8+*	(.4+)	
Fam. Figures	.8+*	(.3-)	
F F x P. Routine		.5+	
Dr-a-M x Int-Attn.		.2-	
Model F	7.54**	2.90**	

*P < .05 **P < .01

( ) indicates first order term not interpretable since the variable is a member of a second order term

[1] N = 22 classrooms

[2] N = 503 pupils

[3] For comparability with Between Class variance, multiply each percent by 13.5.

Table 4.20

Variance in Copying Accounted for,
First Grade

Measures	Models	
Between Classrooms[1]	9 Variable	3 Variable
P12 Differ. Activ.	17-*	
P15 Interest-Attention	9-	32-**
P19 T Strong, Negative	7+	45-**
Mean Draw-a-Man	0-	8-
Mean Fam. Figures	8+	
Mean SES	10+	
Dr-a-M x Diff. Act.	18+*	
SES x T Strong, Neg.	12-*	
Diff. Act.[2]	10+	
Model F	3.79*	5.26**
Within Classrooms[2]	4 Variable[3]	2 Variable[3]
Pretest		.5-
Draw-a-Man	.3+	1.4+**
Fam. Figures	.2-	
F F x P. Routine	.2+	
Dr-a-M x P. Guess	.2-	
Model F	1.79	3.52*

*P<.05 **P < .01

[1]N + 22 classrooms

[2]N = 503 pupils

[3]For comparability with Between Class variance, multiply each percent by 4.1.

( ) First Order term cannot be interpreted since the variable is a member of a second order term

the three variable model appears to be preferable to the nine variable one. It is significant at the one percent level rather than the five, and accounts for almost as much variance with three variables as the other does with nine. The significant relations are for Pupil Interest-Attention and for Teacher Strong Control and Negative Affect. The latter relation agrees with expectations, but the relation for Interest-Attention does not. It suggests that pupils gained less in a relatively complex outcome measure (which is similar to an item on the Stanford-Binet) in classrooms where larger numbers of pupils are judged to be attending to task. The parallel of this relation to the one for Numbers, which approached significance, offers additional, although minor support for the interpretation that it may be possible to focus a classroom too closely for most gain in complex outcomes.

Within Classroom, the two variable model was the only one which was significant. It showed the usual negative relation for pretest, and a small positive one for Draw-a-Man. Both relations parallel findings for previous outcome measures.

Summary

A summary of the models which seemed preferable is presented in Table 4.21 to facilitate identification of pupil characteristics and classroom process measures which were related to more than one outcome. Between Classes, Mean Draw-a-Man was significantly related for Word Meaning and Numbers in the expected direction, but was negatively, but nonsignificantly, related with Copying. SES was related with Listening and Numbers in a direction which indicated greater learning for low SES pupils, presumably because it had earlier been partialled out of the regressed gain scores.

Of the process measures, Differentiated Activities was related negatively with Listening and nonsignificantly with Numbers. This finding agrees with the Medley review (1977) which questions the usefulness of small group activities. The items in this measure reflected considerable pupil freedom, along with the multiple activities and it seems likely that this may have accounted for the negative relation with gain. Pupil Interest and Attention to Task was negatively related to gain in two complex outcome measures, although one relationship only approached significance, suggesting that classroom focus may be too great for learning in some complex outcomes. This finding appears to raise a qualification for the recommendation of direct instruction for all classrooms and all outcome measures. The strong negative relationship between Teacher Strong Control and Negative Affect and pupil gain follows expectation; perhaps the surprise is that it only appears for two outcome measures.

For the Within Class analyses, probably the most important result is the relatively large proportions of variance accounted for when the comparability multiplier is applied. But within these analyses most variance was accounted for by pupil characteristics, with interactions with classroom process accounting for relatively small amounts of

Table 4.21

Summary of Variance Accounted for by Preferred Models, First Grade

Variable	Word Meaning	Listening	Matching	Numbers	Copying
Between Classes					
First Order Variables					
Mean Pretest		4-			
Mean Draw-a-Man	39+**			23+*	8-
Familiar Figure		15+*			
SES (1=High;7=High)		9+		16+*	
P21 Positive Affect	28-**				
P6 Gentle to Mod. Cont.			24-*		
P11 Teacher Central	16-*				
P12 Different'd Activ.		14-*		4-	
P15 Interest-Attn.				10-	32-**
P 19 T Strong Control,Neg.		50-**			45-**
N of Variables in Model	3	5	1	4	3
Within Classes					
First Order Variables					
Pre Alphabet	1+*	4.5+**	1.8+**	2.8+**	
Pretest		1.8-**	2.8-**	3.8-**	.5-
Draw-a-Man				.8+*	1.4+**
Familiar Figures	(1+*)	.5+	2.0+**	.8+*	
Race		(1.5-**)			
Second Order Variables					
P12 X FFT	1.3-*				
P19 X Race		1.2+*			
N of Variables in Model	3	5	3	4	2
Comparability Multiplier	11.2	8.7	10.4	13.5	4.1

*P < .05; **P < .01; ( ) not interpretable, since the variable is a member of a second order term.

variance. There is modest support in these data for the hope that the interaction of classroom process with pupil characteristics within classrooms will produce further understanding of effective teaching.

Among the pupil characteristics which were related to gain, Draw-a-Man showed moderate relationships as it had between classrooms. These data were collected during the first year of a court-ordered bussing plan which mandated equal racial proportions for schools so that classrooms were probably more nearly alike than would typically be true.

The Alphabet measure which was included in the predictors with the hope that it would reflect support for learning in the home was related positively with four of the five outcomes. Familiar Figures, which had related with pupils' learning to read in its development, was related significantly positively with three of the outcome measures. The relations for pretest and race were interpreted as spurious and due to previous adjustment of classroom level effects after having been partialled out in the estimation of individual gain measures.

There were only two interactions between pupil characteristics and classroom process, within class, accounting for modest amounts of variance. One interaction indicated that reflective pupils gained more in classrooms where differentiation was low; the other indicated that white pupils gained more than black pupils where Teacher Strong Control and Negative Affect was high.

In comparing the results for the Between and Within Class analyses, it should be remembered that the Between Class percentages are based on classroom level variance, as usually calculated, which itself is a small percentage of total variance. Since the total variance was the base for the Within Class percentages, they are on a quite different scale. The "comparability multiplier " permits comparing them.

The major conclusions, then, from the first grade Within Classroom results appears to be that the variables account for amounts of variance comparable with the Between Classroom analyses, but that of that variance, most is accounted for by pupil characteristics. There is only modest encouragement here that interactions between pupil characteristics and classroom behavior will clarify the nature of effective teaching.

Process-Product Analyses for Fifth Grade

Screening

As noted in the Procedure Chapter and summarized for first grade, a first step in studying process-product relations was a step-wise multiple regression as a rough screening analysis for both Between and Within Class relations. The variables screened for the Between Class analyses are shown in Table 4.22. A total of 46 variables was screened under the rationale cited earlier and with the qualification also cited there. For the Within Class analyses, the variables screened are shown in Table 4.23; although 57 variables were screened, this number is small in relation to the degrees of freedom.

For Vocabulary and Spelling, 59 classrooms were available for analysis, but because Reading and Arithmetic were sometimes taught in classrooms other than those observed, there were 52 classrooms for Reading and 53 for Arithmetic. In addition to differences in numbers of pupils from subject-matter to subject-matter, a maximum of 42 pupils who were missing SES data were omitted from analyses involving individual SES data, so the pupil N varied from 1085 to 1272.

Models

Vocabulary. The models tested for Vocabulary gain Between Classes are shown in Table 4.24. This table is interpreted in the same way as the first grade results, discussed on pp. 42-44. Again, the attempt was made to provide both first and second order models for each outcome.

For Vocabulary, probably the ten variable model would be preferred. For the two briefer models, which contained only first order terms, one had only one variable significant, the other none, although both models were significant. Within the ten variable model, Teacher and Pupil Positive Affect was negatively related to gain, as it was for Word Meaning in first grade. For the first grade results the explanation was suggested that this measure probably reflected freedom of behavior for pupils but did not reflect task involvement. However, for fifth grade, Pupil Interest-Attention was part of each model, and was partialled out of the relation for positive affect. It was a rough measure of task involvement, but task was not necessarily subject-matter. However, the data were collected in 1970, in a relatively traditional area where open classrooms had not been adopted, and approximately half of the classrooms were in rural areas. It seems likely that the overlap between task and subject-matter was probably extensive.

An examination of the items entering the two original measures of teacher and pupil positive affect (pp.28-29), suggests that all of them represent behaviors which educators would hold in high regard. But these results for fifth grade, and those for first grade Word Meaning, suggest the possibility that the educator's regard may be partially misplaced. Perhaps, instead, pupil gain is greater in classrooms in which approval is given on the basis of performance rather than given unqualifiedly.

Table 4.22

Variables Screened, Between Classrooms
Fifth Grade[1]

Measure	First Order	Sq	Interaction With									
			Pdm	IQ	SES	Sex	Race	Pretest	Anx	Dep	Int	Mot
Paradigm Factor												
P6 T Gentle to Moderate Control	X	X	15									
P8 Pupils Follow Routine	X								X		X	X
P11 Teacher Central	X	X	12							X		
P12 Differentiated Activities	X	X	15		X							X
P15 Interest-Attention Rating	X											
P16 T Guided Discovery (Facts)	X	X			X			X				
P17 Narrow, One-Answer	X	X			X			X				
P18 Pupils Guess	X											
P9 Low Cognitive Level	X	X						X				
P10 Mid to High Cognitive Level	X											
Pooled												
P19 T Strong Control & Neg. Aff.	X		6		X			X	X			
P20 P. Negative & Disorder	X											
P21 T & P Positive Affect	X											
Pupil Characteristic--class mean												
IQ	X											
SES	X											
Sex	X											
Race	X											
Pretest Reading	X											
Anxiety	X											
Dependency	X											
Internality	X											
Motivation	X											

[1]N = 52 to 59 classrooms

Table 4.23

Variables Screened, Within Classrooms
Fifth Grade[1]

Measure	First Order	Sq	Interaction With: Pdm	IQ	SES	Sex	Race	PreRead	Anx	Dep	Int	Mot
Paradigm Factor												
P6 T Gentle to Moderate Control								X	X	X		
P8 Pupils Follow Routine						X			X	X	X	X
P11 Teacher Central										X	X	
P12 Differentiated Activities				X		X					X	X
P15 Interest-Attention Rating					X	X			X	X	X	X
P16 T Guided Discovery (Facts)				X	X							
P17 Narrow, One-Answer								X		X	X	
P18 Pupils Guess				X								
P9 Low Cognitive Level								X	X			
P10 Mid to High Cognitive Level				X								
Pooled												
P19 T Strong Control & Neg. Aff.					X		X	X	X			
P20 P. Negative & Disorder					X	X		X	X	X		
P21 T & P Positive Affect					X							
Pupil Characteristics												
IQ	X											
SES	X	X										
Sex	X											
Race	X											
Pre Voc	X											
Reading	X	X										
Spelling	X											
Arithmetic	X											
Anxiety	X	X										
Dependency	X	X										
Internality	X	X										
Motivation	X	X										

[1]N = 1085 to 1272 pupils

Table 4.24

Variance in Vocabulary Accounted For,
Fifth Grade

Measures	Models		
Between Class[1]	10 Variable	5 Variable	4 Variable
Mean Sex (1-male;2-female)	7+*	4+	
Mean Reading	7-**		
Mean Motivation	14+**	5+	4+
P6 Gentle to Mod Control	10-**		
P12 Differentiated Activ.	9+**	5+	6+
P15 Interest-Atten.	6-*	5+	4+
P21 T & P Positive Aff.	9-**	7-*	6-
Motiv. x Differ. Activ.	16-**		
Differ. Activ. x Int-Att.	9+**		
Reading x Gentle to Mod. Cont.	12+**		
Model F	4.80**	10.41**	2.95*
Within Class[2]	7 Variable[3]	3 Variable[3]	2 Variable[3]
PreReading		1.4+**	1.0+**
Anxiety	.7-**	.7-**	.6-**
Externality	.4+*		
SES (1-High;7-Low)	1.0-**		
Sex (1-male;2-female)	.5+**		
Race (1-black;2-white)		.9-**	
Exter x Int-Attn.	.5-*		
Sex x P. Neg. Disorder	.6-**		
SES x P. Neg. Disorder	.9+**		
Model F	5.25**	11.74**	11.92**

[1] N = 59 classrooms

[2] N = 1230 to 1272 pupils, depending on the model.

*p < .05 **p < .01

(oval) Not interpretable, since the variable is a member of a second order term.

[3] For comparability with Between Class variance, multiply each percent by 8.6.

The findings for praise (an aspect of positive affect, and an item in this factor) (Rowe,1974; Brophy and Evertson, 1974) suggest the negative consequences cited earlier, but in addition Rowe notes that teachers praise low performing pupils about as much as they do high performing pupils, but do so on an essentially random basis. Praise (and perhaps positive affect generally) given in this manner could well have a negative effect on performance.

In general, these findings as a group would seem to raise question about the "happy classroom" as a necessary condition for academic learning. On the other hand, they do not imply the desirability of a negative, critical classroom; the results also indicate that it is counterproductive. Clearly we need to know more about the context and use of positive affect in the classroom in order to clarify this relationship.

Mean sex was related, but since it had been adjusted out of the individual regressed gain scores, probably this result reflects a difference in level of aggregation. The interaction between Motivation and Differentiated Activities indicated that for classrooms with little differentiation, pupil motivation made little difference, but for highly differentiated classrooms, low motivated pupils learned more than high (Figure 4.3). Although it goes beyond the data, perhaps a classroom with multiple activities helped to arouse interest. Again, although it is speculative, perhaps these results parallel those for Listening in first grade in which reflective pupils learned more than impulsive pupils in classrooms which were little differentiated. Perhaps the common element is the willingness of the pupil to work at the task, considering the reflective pupil(perhaps like the highly motivated) as the one who is willing to spend time and effort comparing highly similar figures. Perhaps task differentiation arouses interest for the low motivated pupil, but distracts the pupil who is ready to work.

The interaction of Differentiated Activities with Interest-Attention (Figure 4.4) indicated that where Differentiated Activities were high, gain was greater for classrooms where pupil interest and attention to task was also high than for classrooms where interest and attention was low, although this difference was small. In the first grade results, Differentiated Activities was negatively related to gain in Listening. That finding was interpreted as a reflection of considerable pupil freedom and multiple activities which might have little task involvement. This interaction for fifth grade suggests that differentiation of activities was more functional for pupil learning if the teacher's management system was able to maintain high pupil task involvement.

The interaction between Reading and Gentle to Moderate Control (Figure 4.5) indicated that for classrooms where Gentle to Moderate Control was high, gain in Vocabulary was greater for pupils high in pretest Reading than pupils whose pretest Reading was low.

For the Within Classroom results, probably the three variable model would be preferable since it accounted for almost as much variance as the 7-variable model, but without the complexity of interactions. Pretest Reading was positively related with Vocabulary gain within classroom, and

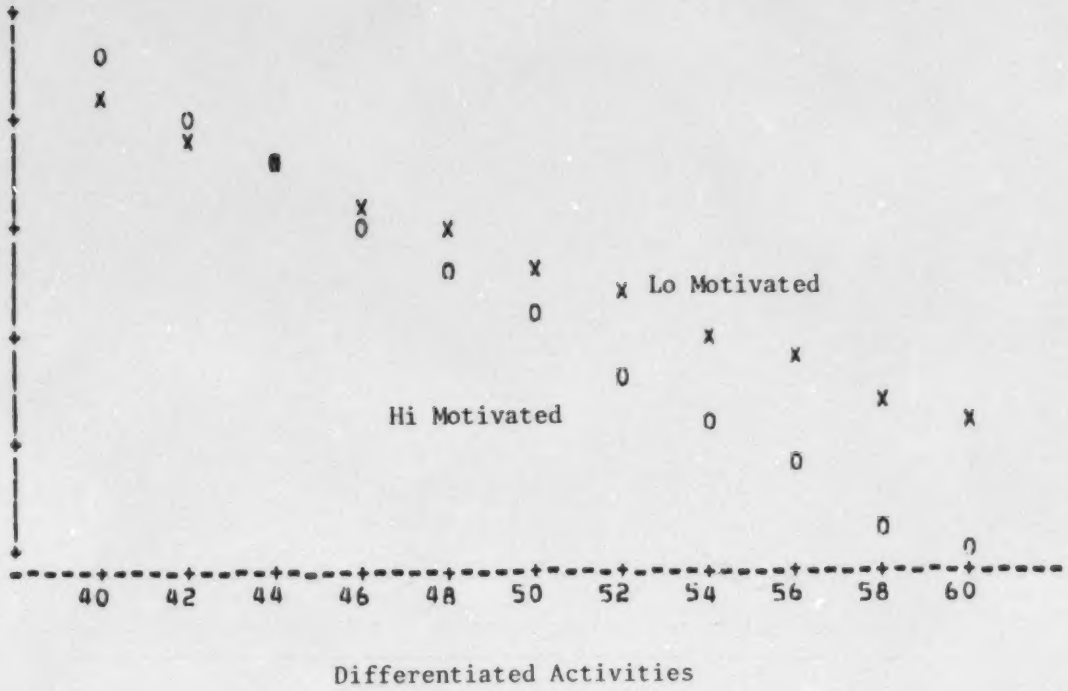


Figure 4.3 Interaction of Differentiated Activities with Motivation for Vocabulary Gain, Between Class, for Fifth Grade

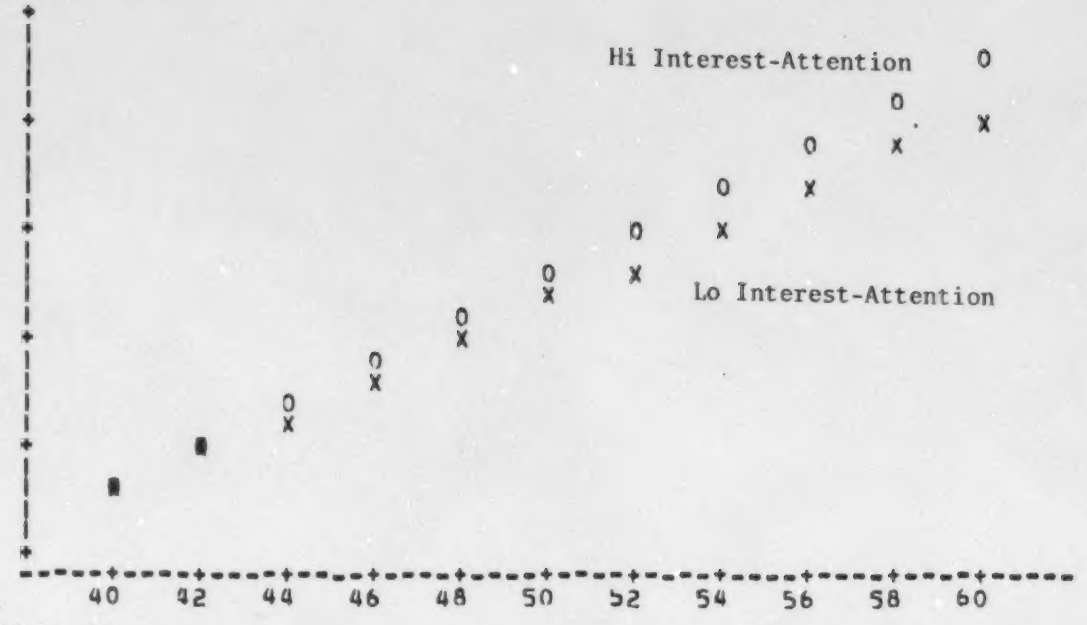


Figure 4.4 Interaction of Differentiated Activities with Interest-Attention for Vocabulary Gain, Between Class, for Fifth Grade

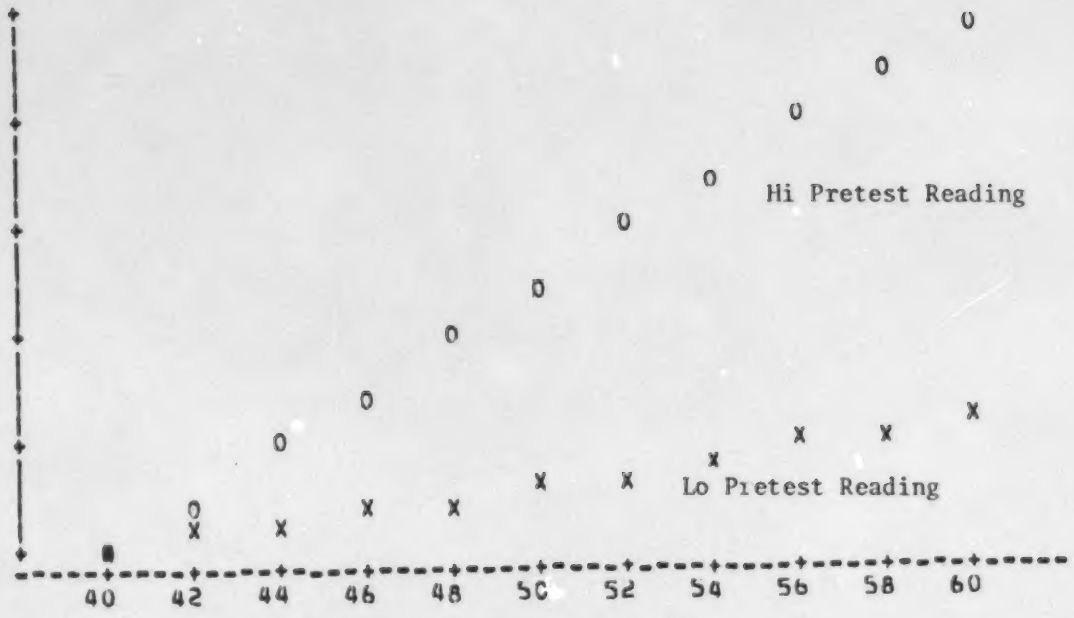


Figure 4.5 Interaction of Teacher Gentle to Moderate Control with Pretest Reading for Vocabulary Gain, Between Class, for Fifth Grade

pretest Anxiety was negatively related. Since Race had been adjusted out of the gain measure, its relation was presumably due to previous adjustment at the classroom level and is not interpreted here.

In general, the findings for Vocabulary differed from those for first grade outcomes in that variance accounted for Within Class was modest compared to that Between Classes, but similar in that the Within Class variance was accounted for primarily by pupil characteristics rather than interactions with classroom behavior. Contrary to first grade, interactions with classroom behavior did appear between classrooms and were relatively strong.

Reading. The results for Reading are presented in Table 4.25. Between Classes, the four variable, first order model suggested that mean classroom Anxiety was a relatively strong predictor of gain and was negatively related as would be expected. The process of teaching reading takes place in a very public setting in the classroom, so perhaps it is reasonable that a high level of anxiety would interfere with learning to read. Pupil Negative Affect and Disorder was also negatively related, as would be expected.

The positive relationship between pupil gain in reading and the amount of classroom interaction which occurred at middle to high cognitive levels is noteworthy. Several studies (Medley, 1977) have shown negative relationships between this style of interaction and gain, but they were primarily studies of first grade, disadvantaged pupils. This result suggests that higher level cognitive interaction may relate positively to gain, as educators expect, at higher grade and/or ability levels.

Between the other two models, probably the nine variable model would be chosen since it accounted for a greater proportion of variance. Its only interpretable first order terms were those of Mid to High Cognitive Level Interaction and class mean anxiety, whose relations were the same as in the first order model just discussed. The positive relation for Mid to High Cognitive Level Interaction was supported by a negative relationship for Narrow, One-Answer interaction, which was not significant, but interesting because it was drawn from a different observation instrument than the data for cognitive level.

The first of the significant interactions (Figure 4.6) indicated that for classrooms where pupils followed a routine to a greater degree, internal pupils learned more than external pupils; for classrooms where there was little routine, external pupils learned more than internal. This finding suggests that where pupils were on their own to follow preestablished procedures, pupils who were inner-directed gained more in reading than pupils who were not.

The other interaction (Figure 4.7) indicated that for classrooms where Teacher Gentle to Moderate Control was high, greater gain took place if Pupil Interest-Attention to task was high than if it was low.

Table 4.25

Variance in Reading Accounted For,
Fifth Grade

Measures	Models		
Between Class[1]	9 Variable	6 Variable	4 Variable
Mean Anxiety	6-*		17-**
Mean Externality	11+**	10+**	
P6 Gentle to Mod. Cont.	10-**	14-**	
P8 P Routine	11+**	10+**	
P10 Mid to High Cog. Level	6+**		7+*
P15 Int-Attn.	4-*	5-*	
P17 Narrow, One-Answer	3-		5-
P20 P Negative, Disorder			8-*
Exter. x P Routine	13-**	13-**	
Int. Att. x Gentle to Mod. Cont.	9+*	13+**	
Model F	8.10**	7.80**	6.19**
Within Class[2]	6 Variable[3]	5 Variable[3]	3 Variable[3]
Pretest Reading	.7-**	1.0-**	.4-*
Anxiety	.0+	.0+	
Dependency		.4+*	.7+**
SES (1-High;7-Low)	.1-		
Race (1-Black;2-White)			.9-**
SES x P Negative Disorder	.1+		
Anx. x P Negative, Disorder	.5+*	.4+*	
Anx. x T Strong, Negative	.7-**	.7-**	
Model F	3.20**	4.55**	7.37**

[1] N = 52 classrooms

[2] N = 1085 to 1120 pupils, depending on model

* $p < .05$ ** $p < .01$

[3] For comparability with Between Class variance, multiply each percent by 7.7.

◯ Not interpretable, since the variable is a member of a second order term.

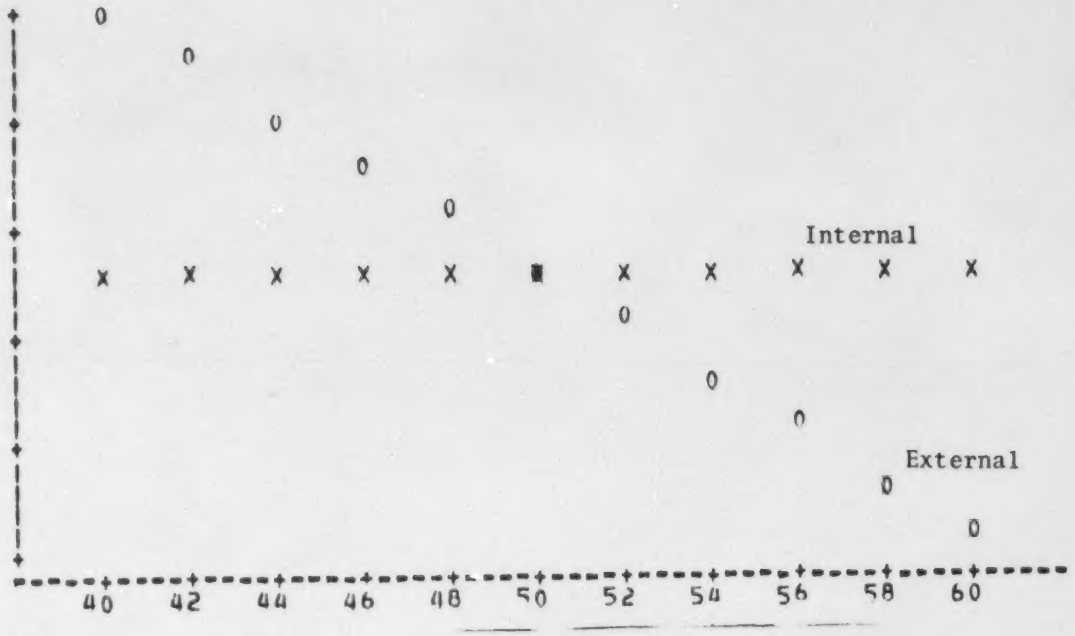


Figure 4.6 Interaction of Pupils Follow Routine with Externality for Reading Gain, Between Class, for Fifth Grade

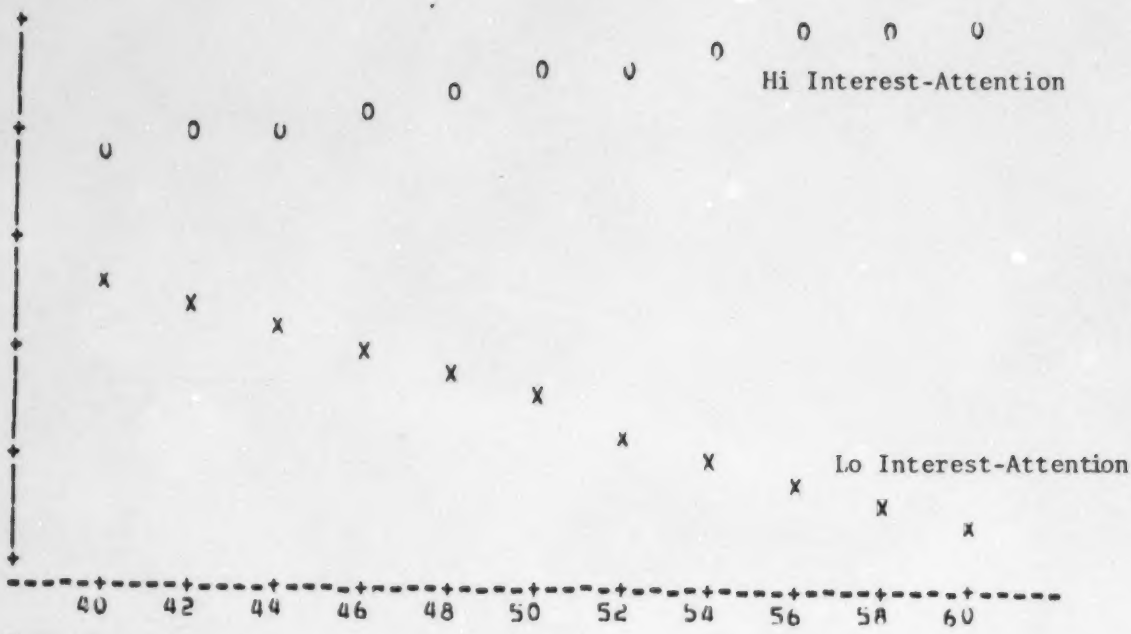


Figure 4.7 Interaction of Teacher Gentle to Moderate Control with Interest-Attention for Reading Gain, Between Class, for Fifth Grade

For the Within-Class analysis the three variable model would be preferable, since it accounted for almost as much variance as the larger, more complex ones. It indicated a positive relation between dependency and gain in reading. The relations for race and pretest are assumed to be uninterpretable since they were previously adjusted out of the gain measures, and adjusted again at the classroom level.

The comparison of Between and Within Class analyses indicates that the Within Class analysis accounts for modest amounts of variance, and that much of what it accounts for is based on pupil characteristics. Again, the meaningful interactions with classroom behavior occurred Between classrooms.

Spelling. The results for Spelling are reported in Table 4.26. For the Between Class analysis, the three variable model appears to be preferable since the other models do not increase the number of variables which are significant. The major influence is for Pupil Interest and Attention to Task, which was positively related to gain. This relation for an outcome measure judged to be simple or low cognitive level was opposite in direction to the two negative relationships which were found for outcomes judged to be complex in first grade--Copying, which was significant, and Numbers, which approached significance.

Gain in Spelling was also positively related to a Narrow, One-Answer style of interaction, as would be expected for a simple outcome. This result contrasted with that for Reading, in which there was a positive relation with higher cognitive level interaction, and a negative (near significant) relation for Narrow, One-Answer style. Again, Reading was an outcome which was judged to be more complex.

For the Within Class analysis, probably the four variable model is preferable, although it adds two interactions which contribute only modest amounts of variance accounted for. In that model the significant variance was accounted for by an interaction (Figure 4.8) which indicated that in classrooms which were not characterized by pupils following a routine, dependent pupils gained more in spelling than independent pupils did, when routine was defined in terms of pupil activities established earlier by the teacher but proceeding largely without current teacher direction. Perhaps in the absence of routine, independent pupils take advantage of the lack of structure and don't work on an uninteresting task such as spelling.

Again, the Within Class analysis accounted for modest variance compared with the Between Class, although in this case that variance was an interaction of a pupil characteristic with a classroom process.

Arithmetic Concepts. The Results for Arithmetic Concepts are reported in Table 4.27. For the Between Class analysis, the variance accounted for was relatively small compared with that for other outcomes. However, an earlier analysis based on process factors (Soar and Soar, 1973) had failed to find any significant relationships. These results indicated that pupils with higher pretest Reading scores gained more in Arithmetic Concepts during

Table 4.26

Variance in Spelling Accounted For,
Fifth Grade

Measures	Models		
Between Class[1]	5 Variables	4 Variables	3 Variables
P9 Low Cog	5-	4-	
P6 Gentle to Mod. Cont.	2+		
P15 Int-Attn.	26+**	25+**	26+**
P17 Narrow, One-Answer	9+**	11+**	7+*
P21 T & P Positive Aff.	6-*	5-	5-*
Model F	6.04**	7.14**	8.02**
Within Class[2]	4 Variables[3]	2 Variables[3]	
Anxiety	(.2+)	.3-	
Race (1-Black;2-White)		.4-*	
Dependency	(.6+**)		
Dep. x P Routine	.5-**		
Anx x P Routine	.3-*		
Model F	3.26*	3.81*	

[1] N = 59 classrooms

[2] N = 1272 pupils

* $p < .05$ ** $p < .01$

[3] For comparability with Between Class variance, multiply each percent by 8.4.

( ) Not interpretable, since the variable is a member of a second order term.

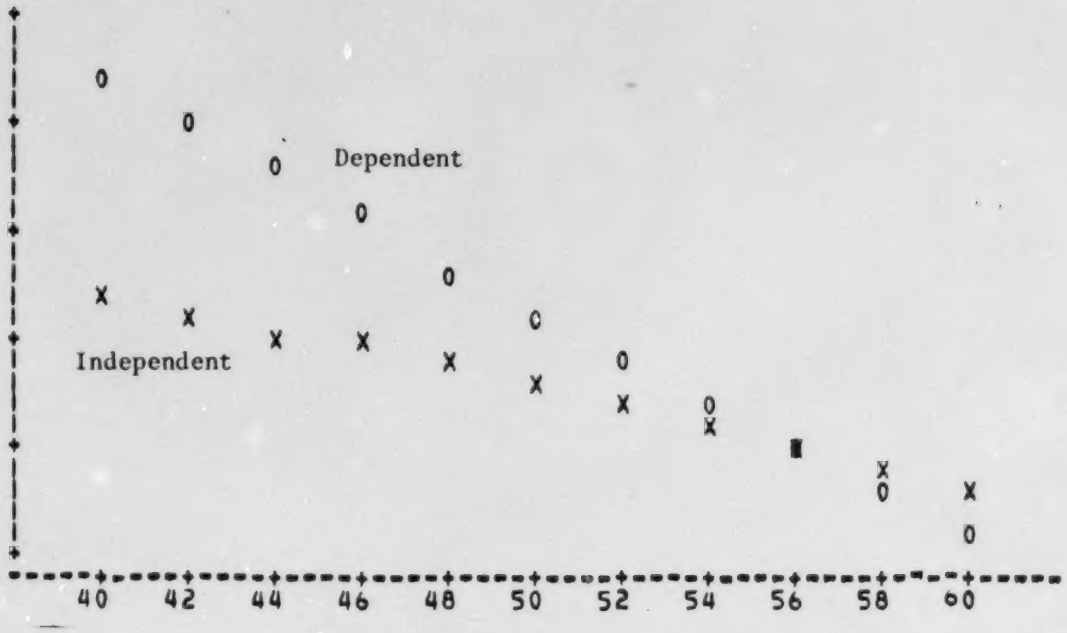


Figure 4.8 Interaction of Pupils Follow Routine with Dependency for Spelling Gain, Within Class, for Fifth Grade

Table 4.27

Variance in Arithmetic Concepts Accounted For,
Fifth Grade

Measures	Models		
Between Class[1]	2 Variable		
Pre Reading	8+**		
P18 P Guess	4-*		
Model F	3.56*		
Within Class[2]	4 Variable	2 Variable	1 Variable
Pre Reading	2.1+**	1.4+**	.8+**
Arithmetic	1.7-**		
Dependency	.6-**		
Lorge Thorndike		.6-**	
Dep. x Narrow, One-Ans.	.8+**		
Model F	9.63**	7.74**	9.30**

[1]N = 53 classrooms

[2]N = 1146 pupils

*p < .05 **p < .01

( ) Not interpretable, since the variable is a member of a second order term.

[3]For comparability with Between Class variance, multiply each percent by 9.2.

the year and that pupils learned less in classrooms in which the interaction was characterized by Guessing which did not involve evaluation of the ideas proposed.

For the Within Class analysis, probably the four variable model would be preferable since it was clearly the strongest in variance accounted for and number of variables which were significant. Pretest Reading was again positively related, even after classroom level effects had been partialled out; and pretest Arithmetic was negatively related--the same negative relationship for pretest which has been noted earlier. There was also a significant interaction indicating that dependent pupils learned more than independent pupils did in classrooms characterized by large amounts of Narrow, One-Answer interaction (Figure 4.9).

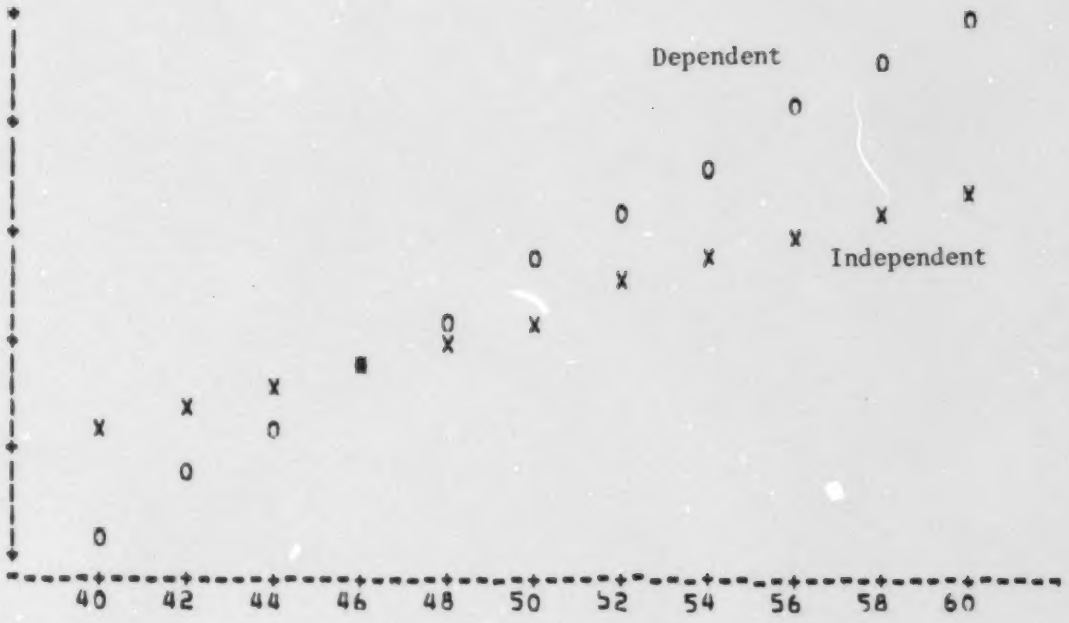
The usual finding of greater variance accounted for by the Between than the Within Class analysis was not obtained for Arithmetic, but it seems probable that the difference is a function of the failure of the process measures to account for much variance in this pupil gain measure, as noted in discussion of the Between Class level analysis.

Summary

The preferred models for the four gain measures for fifth grade are summarized in Table 4.28. Since the preferred models typically contained second order terms, few first order terms were interpretable. In order to permit examination of them, Table 4.29 brings the first order terms together in summary form.

For the Between Class analyses, Anxiety related only to gain in Reading, but that negative relationship was strong. As we suggested earlier, reading in most classrooms is a very public performance, and likely to be stressful for pupils who do not read well, so that a high level of anxiety might be particularly disabling. The other pupil characteristic was pretest Reading, which showed a positive relation with gain in Arithmetic Concepts. It is surprising that it only showed one relationship, but perhaps this is the most reasonable one. It is hard to imagine dealing with concepts in mathematics without reading at least adequately.

Among the classroom behavior measures, Mid to High Cognitive Level Interaction showed a positive relation with Reading gain. This is particularly interesting because of the consistency with which higher cognitive level interaction has related negatively to gain in studies of lower grade disadvantaged pupils, and suggests that the usefulness of this style of interaction may depend at least partly on age or ability. This result contrasted with that for Narrow, One-Answer interaction, which appeared twice, relating positively with gain in Spelling; and negatively, although not significantly, with Reading. There appears to be support here for the idea that the outcome measures differ on some basis. The project proposal hypothesized Spelling to be a simple outcome, and Reading to be a more complex one (with Vocabulary the most complex). The finding that Reading was related with higher level interaction, and spelling with the one-answer style, seems to offer some support for this hypothesis.



Narrow, One-Answer Interaction

Figure 4.9 Interaction of Narrow, One-Answer with Dependency for Arithmetic Concepts Gain, Within Class for Fifth Grade

Table 4.28

Summary of Variance Accounted for by Preferred Models, Fifth Grade

Variable	Vocabularly	Reading	Spelling	Arith. Con.
Between Class				
First Order Variables				
Mean Pretest Reading	7-**			8+**
Anxiety		6-*		
Motivation	14+**			
Externality		11+**		
Sex (1-male;2-female)	7+*			
P6 Gentle to Mod. Cont.	10-**	10-**		
P8 P. Routine		11+**		
P10 Mid to High Cog.		6+*		
P12 Different'd Activ.	9+**			
P15 Interest-Attn.	6-*	4-*	26+**	
P17 Narrow, One-answer		3-	7+*	
P18 P. Guess				4-*
P21 T & P Positive Affect	9-**		5-*	
Second Order Variables				
Motiv. x Differ. Activ.	16-**			
Differ. Activ. x Int-Attn.	9+**			
Reading x Gentle to Mod.Cont.	12+**			
Externality x P. Routine		13-**		
Int-Atten. x Gentle to Mod.Cont.		9+*		
No. of Variables in Model	10	9	3	2
Within Class				
First Order Variables				
Pre Reading	1.4+**	.4-*		2.1+**
Arith. Con.				1.7-**
Anxiety	.7-**		.2+	
Dependency		.7+**	.6+**	.6-**
Race	.9-**	.9-**		
Second Order Variables				
Dependency x P. Routine			.5-**	
Anxiety x P. Routine			.3-*	
Dependency x Narrow, One-Ans.				.8+**
No. of Variables in Model	3	3	4	4
Comparability Multiplier	8.6	7.7	8.4	9.2

*p < .05 **p < .01 (oval) Not interpretable, since the variable is a member of a second order term.

Table 4.29

Summary of Variance Accounted for by
First Order Models, Fiftn Grade

Variable	Vocabulary	Reading	Spelling	Arithmetic Concepts
Between Class				
PreReading				8+**
Sex (1=male;2=female)	4+			
Motivation	5+			
Anxiety		17-**		
P6 T. Gentl. to Mod. Cont.				
P10 Mid to Hi Cog. Level		7+		
P12 Differ. Activ.	5+			
P15 P Int.-Attn.	5+		26+**	
P17 Narrow, One-Ans.		5-	7+*	
P18 P Guess				4-*
P20 P Neg.,Disorder		8-*		
P21 T & P Pos. Affect	7-*		5-*	
No. of Variables in Model	5	4	3	2
Within Classes				
PreReading	1.0+**	.4-*		.8+**
Anxiety	.6-**		.3-	
Dependency		.7+**		
Race		.9-**	.4-*	
No. of Variables in Model	2	3	2	1
Comparability Multiplier	8.6	7.7	8.4	9.2

*p < .05 **p < .01

The fact that the positive relationship between pupil Interest and Attention to task and gain in Spelling was the strongest relationship in the table may also be related. Perhaps the classroom activity in which task involvement can be most clearly observed is that which supports simple learning outcomes. This strong relation with Spelling in contrast with the others appears to raise question about the belief that obvious pupil task involvement is a generalizable aspect of effective teaching.

Pupil Negative Affect and Disorder was related negatively with gain in Reading. This is the expected direction, but it is surprising that it did not relate to other outcomes.

Teacher and Pupil Positive Affect related negatively with both Vocabulary and Spelling, contrary to educators' beliefs in its value. While the classroom measure may reflect lack of task focus, the possibility should be considered that positive affect can be used in ways that hinder learning, as discussed earlier. It seems clear that we need to know more about the context and use of positive affect in the classroom.

There were no parallels for the interactions except by implication (Table 4.28). For classrooms where Gentle to Moderate Control was high, greater gain in Reading occurred if pupil Interest-Attention was high and greater gain in Vocabulary occurred if pupils were high in pretest Reading. Apparently, Gentle to Moderate Control was a functional style if pupils had the skills to learn or if the teacher had otherwise established involvement with the task.

Two other interactions indicated that classrooms where there was much differentiation of activity supported learning if pupils began the year low in motivation or if the teacher's management system otherwise maintained pupil Interest and Attention to task. A final interaction suggested that classrooms where there was high Pupils Follow a Routine supported learning for pupils who were internal in control.

For the Within Class analyses, pretest Reading related positively to gain in both Vocabulary and Arithmetic Concepts. The negative relation of Pretest Reading to Reading gain is presumably uninterpretable due to the previous adjustment in calculating regressed gain.

The negative relation between Anxiety and Vocabulary gain was not strong, but it is perplexing that Anxiety related with Reading gain at the Between Class level and with Vocabulary gain Within Class.

As discussed in the summary for first grade, it should be remembered that the Between Class results and the Within Class results have different bases for the percents of variance accounted for. The comparability multiplier applied to the Within Class results would have the effect of bringing them to the same base.

In summary, the major variance accounted for appeared for the Between Class analyses, with only modest proportions of variance appearing for the Within Class analyses. Paralleling first grade, relations with classroom behavior were prominent at the Between Class level, but most of the variance which was accounted for at the Within Class level was a function of pretest pupil characteristics. Again, there is only modest support for the hope that recognizing interactions between pupil characteristics and classroom process Within Class will bring a material advance in the understanding of effective teaching.

Process-Product Analyses for First and Fifth Grade

Process Differences Between Grades

Although the major interest in the combined analysis of the first and fifth grades was the question of whether classroom behaviors which were related to gain might differ significantly from one grade level to the other (e.g., might interact with grade level), a preliminary question was whether classroom behavior itself differed from one grade level to the other. In order to answer this question, classroom behavior was compared across the two grade levels using the t test.

The results are shown in Table 4.30. Only one of the measures failed to discriminate significantly between the grade levels--the Interest-Attention rating, for which first and fifth grades means were virtually identical. The differences between means for other measures were not only significant but in many cases were impressively large. With one exception, all of the significant differences were ones in which first grade means were higher than fifth grade means. It seems possible that this consistency of result simply reflects in first grade classrooms a higher general level of activity and diversity of the kind we measured. The one exception was Mid to High Cognitive Level interaction which was greater in fifth grade classrooms. Both the greater age of the pupils and the stronger academic focus of the fifth grade make that seem reasonable.

Screening

As indicated earlier, a preliminary step in the process-product analyses was a step-wise multiple regression for both Between Class and Within Class in order to identify a smaller number of promising variables for further analysis by SAS GLM. For the Between Class analysis,40 Variables were studied, identified in Table 4.31. Since there were 81 classrooms in the combined sample, again this number of variables was too large to permit dependable conclusions to be drawn, but the large number was explored following the rationale cited earlier. For the Within Class analysis, 44 variables were screened, as shown in Table 4.32, but since 1733 pupils were studied,this number of variables would be satisfactory for most methodologists.

Models

It will be recalled that different achievement measures were used at first grade than at fifth grade. In order to carry out combined analyses of the two grades, a measure was selected at each grade level which was judged to be cognitively complex (Vocabulary for fifth, Numbers for first), another which was judged to be simple (Spelling for fifth, Word-Meaning for first), and in addition a Total score was created, made up of four measures at each grade level. Scores were transformed to make grade levels comparable.

Table 4.30

Differences in Classroom Behavior
From First to Fifth Grade

Measure	Mean 1st[1]	Mean 5th[2]	Stan. Dev. 1st	Stan. Dev. 5th	t	p
P6 T Gentle to Moderate Control	57.59	47.12	4.16	4.06	-10.26	.000
P8 Pupils Follow Routine	54.73	48.05	5.14	5.52	-4.93	.000
P9 Low Cognitive Level	58.18	46.97	3.74	5.47	-8.85	.000
P10 Mid to High Cognitive Level	42.68	52.47	5.69	6.41	6.30	.000
P11 Teacher Central	53.59	48.44	2.50	5.71	-4.08	.000
P12 Differentiated Activities	54.91	48.34	3.16	4.74	-6.01	.000
P15 Interest-Attention Rating	50.45	49.51	10.42	9.45	-.39	.698
P16 T Guided Discovery	52.41	49.15	5.56	5.47	-2.37	.020
P17 Narrow, One-Answer	53.41	48.44	5.68	5.77	-3.46	.001
P18 Pupils Guess	54.64	48.49	4.29	4.89	-5.19	.000
P19 T Strong Control & Neg. Aff.	52.23	49.15	6.76	5.88	-2.01	.048
P20 P Negative & Disorder	54.27	48.49	6.41	5.32	-4.11	.000
P21 T & P Positive Affect	107.77	96.98	10.30	11.16	-3.95	.000

[1] N = 22 classrooms

[2] N = 59 classrooms

Table 4.31

Variables Screened, Between Classrooms
First and Fifth Grades[1]

Measure	First Order	Sq	Interaction With Pdm	Pre	Gr.	SES
Paradigm Factor						
P6 T Gentle to Moderate Control	X	X	15		X	X
P8 Pupils Follow Routine	X	X				
P11 Teacher Central	X	X	12			X
P12 Differentiated Activities	X	X	15	X	X	
P15 Interest-Attention Rating	X					X
P16 T Guided Discovery (Facts)	X	X				
P17 Narrow, One-Answer	X	X		X		
P18 Pupils Guess	X					
P9 Low Cognitive Level	X	X			X	X
P10 Mid to High Cognitive Level	X					
Pooled						
P19 T Strong Control & Neg. Aff.	X		6	X	X	
P20 P Negative & Disorder	X					
P21 T & P Positive Affect	X					
Pupil Characteristic--Class Means						
Pretest	X					
Grade	X					
SES	X					
Sex	X					
Race	X					

[1]N = 81 classrooms

Table 4.32

Variables Screened, Within Classrooms
First and Fifth Grades[1]

Measure	First Order	Sq	Interact With Pdm	Pre	SES	Sex	Race	Gr.
Paradigm Factor								
P6 T Gentle to Moderate Control	X			X	X	X	X	
P8 Pupils Follow Routine	X			X	X	X		
P11 Teacher Central	X							
P12 Differentiated Activities	X				X			
P15 Interest-Attention Rating	X				X			
P16 T Guided Discovery (Facts)	X						X	
P17 Narrow, One-Answer	X							
P18 Pupils Guess	X							
P9 Low Cognitive Level	X			X	X		X	
P10 Mid to High Cognitive Level	X			X				
Pooled								
P19 T Strong Control & Neg. Aff.	X			X		X	X	
P20 P Negative & Disorder	X			X	X	X	X	
P21 T & P Positive Affect	X				X			
Pupil Characteristics								
Pretest	X							X
SES	X	X						
Sex	X							
Race	X							
P6 x P19 x SES								
GR x P19 x SES								
GR x P17 x Pre								

[1] N = 1733

Complex. The models tested for the Complex measure are reported in Table 4.33. For the Between Class analyses only the nine variable model was significant as a model. None of the first order terms was both significant and interpretable. Three second order terms were significant--a nonlinear relation, and two interactions.

The nonlinear relation for Differentiated Activities, when plotted, indicated a negative relation throughout the range of the variable, and one which became somewhat more negative at higher levels (Figure 4.10), although the deviation from linearity was not great. Apparently the presence of multiple activities occurring simultaneously in the classroom was counterproductive but became more so at an increasing rate as the activities multiplied. The items in the factor suggest a high level of pupil freedom and diversity of activities, but there is little evidence that the activity had to do with learning. The factor was also more strongly related to measures of positive affect (which related negatively with gain) than to any others. Apparently the two factors describe diffuse, unfocused activities in which little learning occurs.

The first interaction indicated that for classrooms with small amounts of Teacher Strong Control and Negative Affect, pupils gained more in this Complex outcome when there was a high amount of Gentle to Moderate Control (Figure 4.11). If Teacher Strong Control was high, however, the amount of Gentle Control made little difference in gain. To oversimplify, apparently it is important to have one or the other control styles present at a high level in the classroom, for gain in achievement; however, it seems probable that the gentle style would be preferred for the attainment of other objectives. The use of neither (little or no control) may represent abdication of management responsibility by the teacher.

The second interaction, the only interaction with grade level, indicated that for classrooms where Teacher Strong Control and Negative Affect was high, pupils in first grade gained more than pupils in fifth grade (Figure 4.12). It would not be surprising that harsh teacher control would be less effective or coercive after pupils are old enough for peer support to develop. Unifying in opposition, even in passive resistance, would be less likely for young children. In retrospect. the process-product correlations (Appendix A) suggest that the selection of Numbers as a Complex measure for first grade may not have been the best choice.

For the Within Class analyses, probably the four variable model would be preferable, since it accounted for a similar amount of variance with fewer terms. The significant interaction indicated that in classrooms where there was little Pupil Negative Affect and Disorder, adjusted gain scores for high SES pupils were higher than for low SES pupils (Figure 4.13). The implication appears to be that high disorder was more disruptive for high SES pupils than for low (or facilitative for low SES pupils--which would be a perplexing finding). Again, the difference in base for percentages should be remembered and the multiplier used for comparability.

Table 4.33

Variance in the Complex Outcome Accounted for,
First and Fifth Grades

Measures	Models		
Between Class[1]	9 Variable	7 Variable	2 Variable
First Order Variables			
Grade	6+*	5+	
P6 Gentle to Mod. Control	10+**	7+*	
P12 Differentiated Activ.	5+*	1+	2+
P18 Pupils Guess	2-		
P19 T Strong,Negative	8+**	7+*	
P21 T & P Pos. Affect	4-		
Second Order Variables			
Differ. Activ. Sq.	11-**	9-**	7-*
Gentle to Mod. Cont. x T Strong	9-**	7-**	
Grade x T Strong	6-*	5-*	
Model F	2.13*	1.88	2.19
Within Class[2]	10 Variable[3]	4 Variable[3]	
First Order Variables			
SES (1=High;7=Low)	.2-*	.8-**	
Race (1=Black;2=White)	.0+		
Pretest	.3-*	.1-	
Second Order Variables			
SES x P Neg. & Disorder	.6+**	.7+**	
SES x Interest-Attn.	.2+		
SES x T & P Pos. Aff.	.0-		
Pre x Gentle to Mod. Cont.	.3+	.1+	
Gr x Pretest	.2+		
Race x One-Ans.	.1-		
SES x P Routine	.1-		
Model F	2.84**	4.30**	

*p < .05 **p < .01

[1]N = 81 classrooms

[2]N = 1733 pupils

[3]For comparability with Between Class variance, multiply each percent by 9.8.

( ) First order term not interpretable since it was a member of a second order term

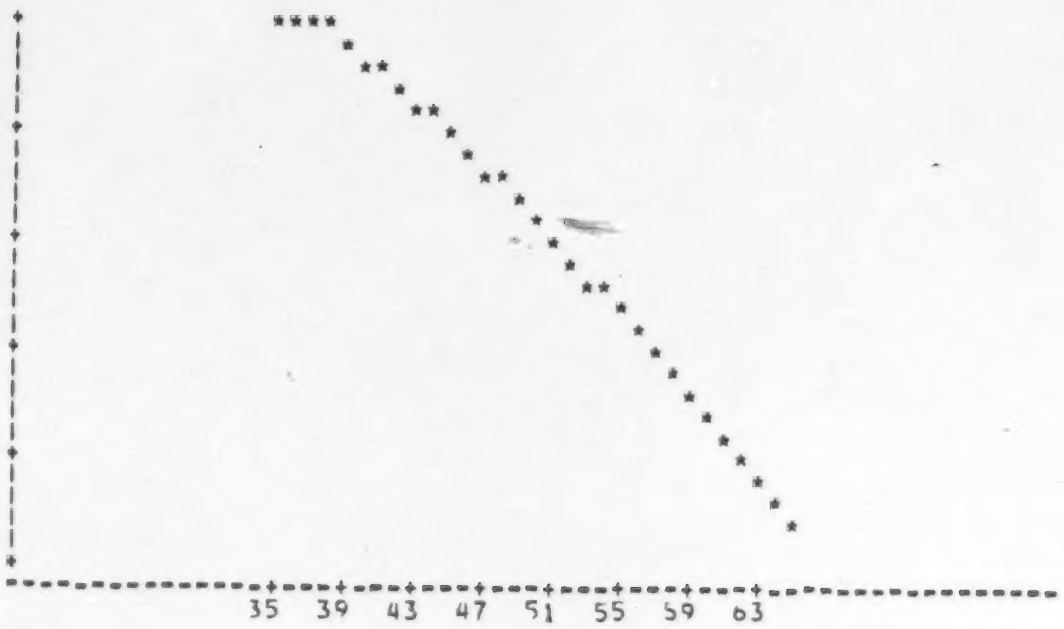


Figure 4.10 Quadratic Relation Between Differentiated Activities and Complex Gain, Between Classes, First and Fifth Grades

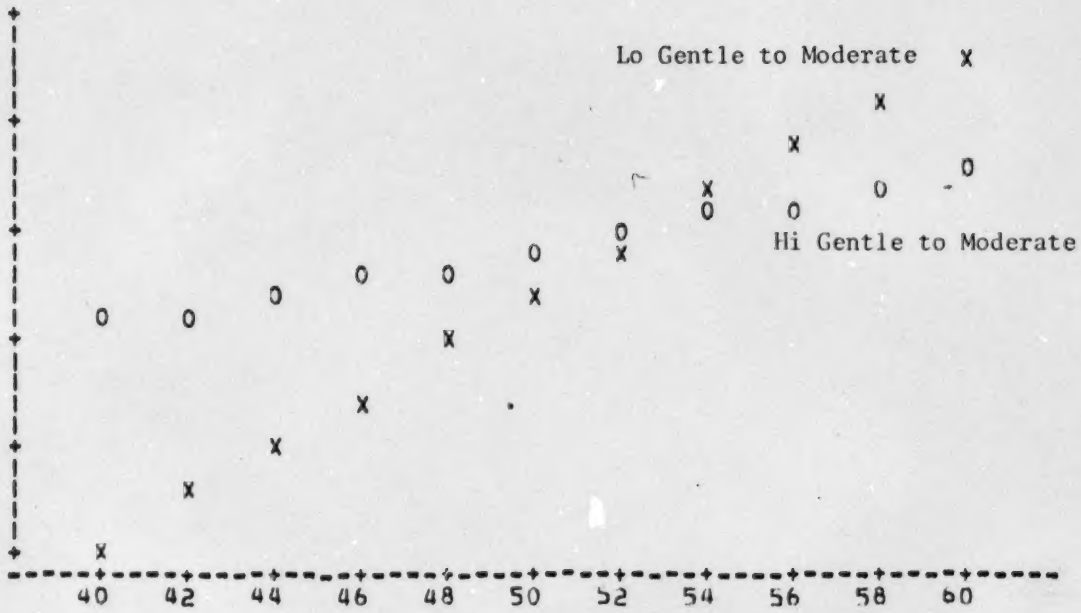


Figure 4.11 Interaction of Teacher Strong Control and Negative Affect with Teacher Gentle to Moderate Control for Complex Gain, Between Class, for First and Fifth Grades

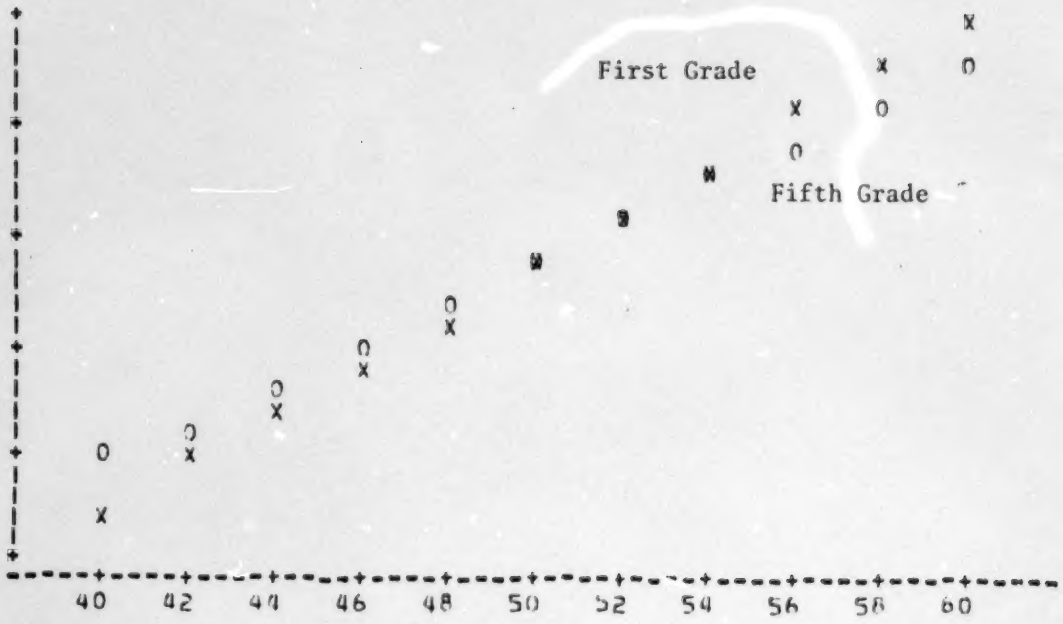


Figure 4.12 Interaction of Teacher Strong Control and Negative Affect with Grade for Complex Gain, Between Class, for First and Fifth Grades

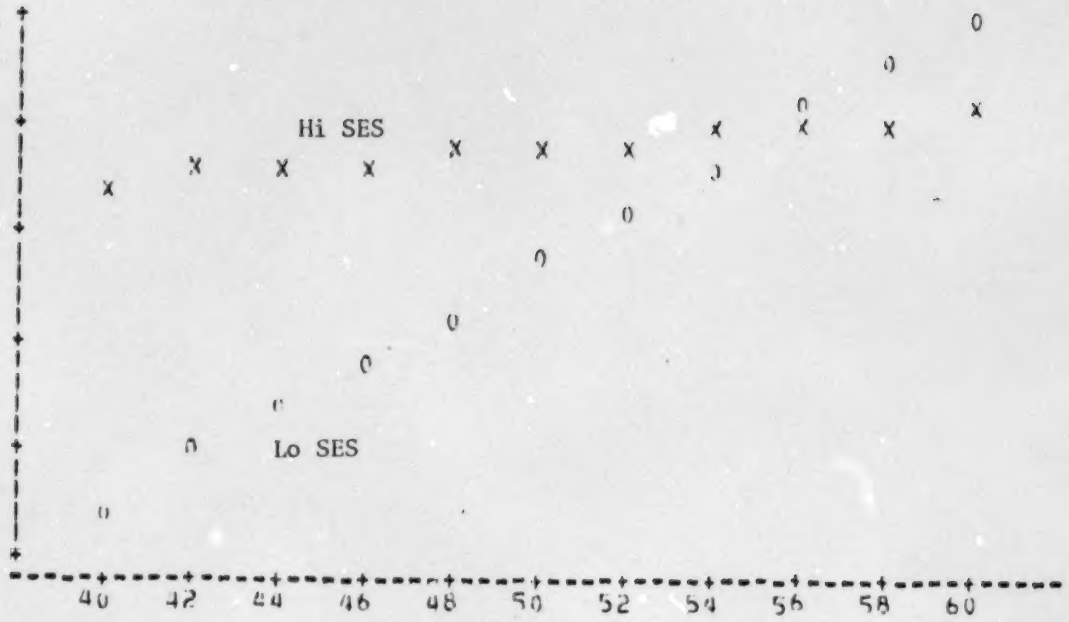


Figure 4.13 Interaction of Pupil Negative Affect and Disorder with SES for Complex Gain, Within Class, for First and Fifth Grades

As in the previous analyses, somewhat greater variance was accounted for in the Between Class analyses than the Within Class analyses. However, the only significant model for Between Classrooms was a very complex one, which, with the inflation due to a large number of variables having been screened, may be suspect. There was one interaction with grade level, a major interest in these analyses, but its effect was small.

Simple. The results for the Simple outcome measure are reported in Table 4.34. For the Between Class analysis, probably the three variable model would be preferred. It accounts for a considerable proportion of the variance of the seven variable model, with fewer than half as many variables and no interactions, and the model F is greater. In addition, it presents a clear picture of effective teaching for a simple outcome measure: high pupil Interest in and Attention to task, a Narrow, One-Answer style of interaction, and little Positive Affect in the classroom. This is a simple, straightforward picture, in contrast to the models for the Complex outcome just discussed and the model for vocabulary at fifth grade.

Although the Within Class analysis resulted in models which were statistically significant, none of the variables accounted for enough variance to warrant interpretation. Again, the selection of measures to represent first grade was not clear-cut, and perhaps not the best.

Total. The results for the measure of Total achievement are shown in Table 4.35. For the fifth grade, the measure of Total Achievement was the sum of Vocabulary, Reading, Spelling and Arithmetic Concepts; for the first grade it was made up of Word Meaning, Matching, Numbers, and Copying.

For the Between Class analysis, if a first order model were to be taken as a description of effective teaching, the three variable model would be preferable to the five variable since all three were individually significant and accounted for as much variance. Pretest and Pupil Interest in and Attention to task were positively related with gain; as before, Teacher and Pupil Positive Affect was negatively related.

Although the 11 variable model accounted for more variance, only one first order term was significant and interpretable, and only one interaction was significant. It seems reasonable, therefore, to accept the three variable model as preferred.

For the Within Class analysis, probably the four variable model would be preferred since it contains as many variables which are individually significant as the ten variable model. It involves a negative relation for Race which is presumably not interpretable because it had previously been adjusted out of the gain measure. There was also a significant interaction indicating that high pretest pupils gained more than low pretest pupils in classrooms where there were large amounts of Mid to High Level interaction (Figure 4.14). It seems reasonable that pupils who have achieved more in the past in a variety of measures should be better prepared to benefit from interaction at a higher cognitive level.

Table 4.34

Variance in the Simple Outcome Accounted for,
First and Fifth Grades

Measure	Model	
Between Class[1]	7 Variable	3 Variable
First Order Variables		
P6 Gentle to Mod. Control	6+*	
P15 Interest-Attention	7+**	13+**
P17 Narrow, One-Answer	4+*	5+*
P19 T Strong, Negative	4+	
P21 T & P Pos. Affect	9-**	8-**
Second Order Variables		
Gent. to Mod. Cont. x T Strong, Neg.	5-*	
Gent. to Mod. Cont. x Int-Atn	6-*	
Model F	3.90**	5.61**
Within Class[2]	6 Variable[3]	4 Variable[3]
First Order Variables		
SES (1=High;7=Low)	.1-	
Race (1=Black;2=White)	.3-*	.2-*
Pretest	.2	.0
Second Order Variables		
Race x P. Neg. Disorder	.2+*	.2+
Pretest[2]	.4+*	.4+**
Pretest x Mid. Cog.	.2+	
Model F	2.79*	3.19*

*p < .05 **p < .01

(6+*) First order term not interpretable since it was a member of a second order term.

[1]N = 81 classrooms

[2]N = 1733 pupils

[3]For comparability with Between Class variance, multiply each percent by 9.3.

Table 4.35

Variance in Total Outcome Accounted for,
First and Fifth Grades

Measure	Model		
Between Class[1]	11 Variable	5 Variable	3 Variable
First Order Variables			
Pretest	2+	6+*	7+*
SES	(2+)		
P6 T Gentle to Mod. Cont.	(6+*)		
P12 Differentiated Activ.	2+		
P15 Interest-Attention	(5+*)	3+	6+*
P17 Narrow, One-Answer	(2+)	2+	
P19 T Strong Control & Neg. Aff.	(5+*)	1-	
P21 T & P Pos. Affect	9-**	5-*	5-*
Second Order Variables			
Gent. Cont. x T. Strong	6-*		
Gent. Cont. x Int-Attn	4-		
SES x Narrow, One-Ans	2-		
Model F	2.83**	3.54**	5.30**
Within Class[2]	10 Variable[3]	4 Variable[3]	
First Order Variables			
SES (1=High;7=Low)	(0.0-)		
Race (1=Black;2=White)	1.8-**	1.8-**	
Pretest	(.2-)	(.0-)	
Second Order Variables			
SES x Gentle Cont.	.0+		
SES x Interest-Attn.	.0+		
SES x T & P Pos. Aff.	.1-		
Pretest x Gentle Cont.	.2+		
Pretest x Mid. Cog.	.7+**	.6+**	
SES x Diff. Activ.	.0+		
Gr x Pre x One-Ans.	.5-**	.9-**	
Model F	6.65**	14.11**	

*p < .05 **p < .01 ( ) First order term not interpretable since it was a member of a second order term

[1]N = 81 classrooms

[2]N = 1733 pupils

[3]For comparability with Between Class variance, multiply each percent by 6.6.

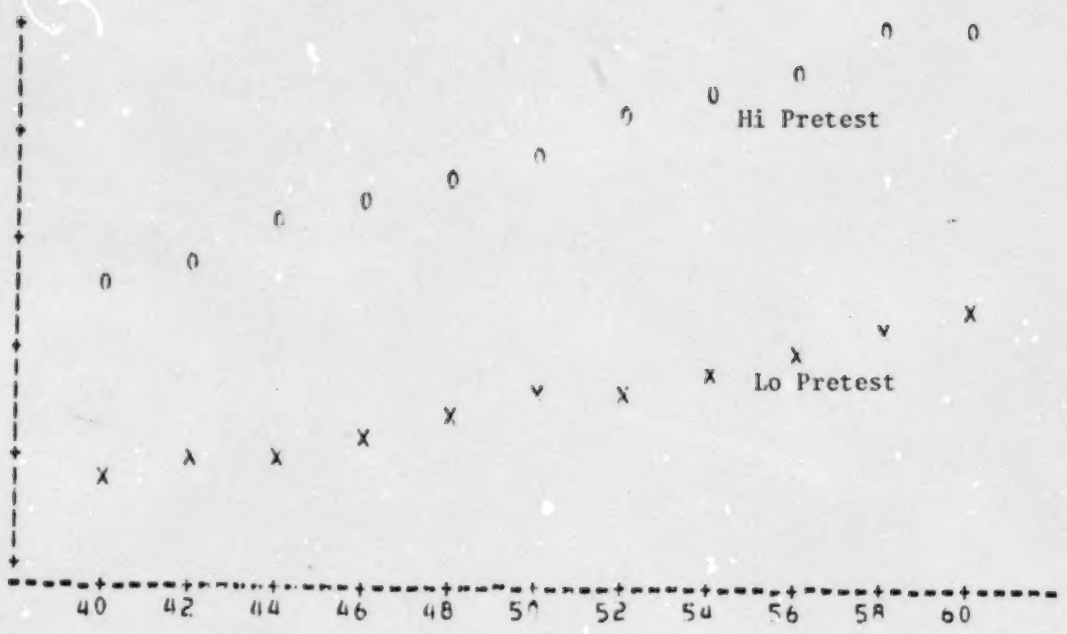


Figure 4.14 Interaction of Mid to High Cognitive Level with Pretest for Total Gain, Within Class, for First and Fifth Grades

The final significant relationship was for a third order variable representing a triple interaction; however, a relationship this complex which accounts for a small amount of variance scarcely warrants interpreting.

Summary

The analysis of data from first and fifth grades was primarily concerned with testing interactions between grade level and classroom process which were most related to gain. However, only one interaction with grade level appeared, and its plot indicated that the differences associated with it were small. Although a clear-cut model emerged describing effective teaching of the Simple outcome measure, a clearer, stronger model had emerged for fifth grade spelling. For the more Complex outcome, only very Complex models were significant and it seemed that the view of effective teaching which emerged from separate analyses of each grade level was clearer and more interpretable.

In addition, the combined analysis may have been weakened not only by the problem of selecting measures at each grade level, but also because the process measures which were related to gain may have differed enough at the two grade levels to weaken pooled relationships, but not enough to create significant interactions.

Chapter 5
Summary and Implications

This project was a reanalysis of data collected earlier, and was planned to address two methodological questions and two substantive questions. The methodological questions were: How can relationships involving pupil characteristics and interactions with classroom behavior be analyzed within classroom, without violating the assumption of independence of observations? and Is there an alternative to traditional methods for forming composite measures of classroom behavior? The substantive questions were: Does the nature of the pupil or the setting make a difference in the teaching style which is most effective? and Does the cognitive level of the learning objective make a difference in the teaching style which is most effective?

The data which were reanalyzed consisted of measures of gain from the Iowa Tests of Basic Skills adjusted for sex, race, IQ and pretest for the total pupil group from 59 fifth grade classrooms, and measures of gain from the Metropolitan Readiness Tests adjusted for sex, race, socioeconomic status and pretest for 22 first grade classrooms. Observation was carried out during the year using four observation instruments (1) a measure of the teacher's management style, pupil's response to it, and expression of affect, (2) a measure of classroom behaviors reflecting agreement or disagreement with Deweyian experimentalism, (3) an extension of the Flanders system which reflects both social-emotional and substantive interactions for both teacher and pupils, and (4) a measure of the cognitive level of teacher-pupil interaction.

Forming Composites for Classroom Process

Two approaches were explored for forming composites for classroom behavior. One approach explored relationships between subsets of items within factors from previous analysis. Several factors seemed subjectively to contain from two to four dimensions with different meanings, which made interpretation difficult, and whose interrelationships were low enough to warrant separate use. However, the total number of measures which resulted was sharply increased, and subfactors from different factors were sometimes so similar as to cause interpretive problems. Consequently that approach was abandoned.

The second approach which was explored started from a paradigm which had emerged from a rational analysis of process factors and process-product relationships from five previous data sets of the authors, as well as consideration of results from the literature in the field. The first step was to identify sets of items from the four instruments which seemed rationally to represent each of the domains specified by the paradigm: emotional climate,which was divided into teacher and pupil, positive and negative; and teacher management, which was divided into management of pupil behavior, of learning tasks, and of thinking. Management of behavior referred to the nonsubstantive activity of pupils--the limits set for physical movement, subgrouping, and socializing. Management of learning tasks reflected teacher direction of the choice and conduct of the learning

tasks in which pupils were engaged; and management of thinking referred to limits set by the teacher for the thinking processes and within those limits the cognitive levels of the interaction which occurred.

The item set for each of these domains was then factored to ensure homogeneity of measures, with the result that each domain broke into at least two factors and one into three. When the resulting factors were intercorrelated, several factors dealing with affect or strong control were sufficiently highly related and similar in meaning, even though they crossed domains, as not to justify separate use, so they were pooled. This process resulted in 12 factors based on the rationale of the paradigm, which in the main preserved its rational structure. In addition, a measure of Pupil Interest and Attention to Task was carried forward as being of interest in its own right and as a measure of current concern in other studies.

The resulting measures seemed intuitively attractive. Not only is a theroetical view of teaching represented in empirical measures, but the univocality of each was tested in its development, and distinctions were sometimes made which were not originally perceived. For example, the Narrow, One-Answer style of interaction sounded highly similar to the measure of low cognitive level interaction, yet the two showed only about a third of their variance in common. In retrospect it is clear that single correct answers can occur at a range of cognitive levels. The distinction between Guided Discovery and Pupils Guess seems obvious, after the fact, but was not beforehand (the titles sharpen the difference), and we suspect the distinction is not often made in the field. Three of the four affect measures (T negative, P negative, and T & P Positive) were found to be relatively independent, and that independence is frequently violated in the use of bipolar, negative-positive rating scales.

We concluded that the paradigm factors were clearer in interpretation than some of the factors previously derived from this data set. For example, a factor containing positive affect and freedom of pupil movement had been negatively related to achievement gain. The interpretation which had seemed reasonable was that the high level of physical activity overwhelmed the return from the positive affect. But the positive affect par digm measure, which related more consistently negatively with gain for both first and fifth grades than any other measure, raises the question of whether positive affect, in itself, may have undesirable consequences. Probably the context and use of the positive affect are important in determining its effect on pupil learning, and need to be further studied. It does seem clear that the "happy classroom" is not only not necessary but may be a liability for pupil gain in achievement, and that was not clear from the previous analysis using conventionally derived factors.

Method of Analysis

Development of an Analysis for Within Class Relationships

After a time spent in the development of a method of analysis which seemed appropriate, we realized that what we had developed was an application (or perhaps an extension) of the split-plot design. At about the

same time, the index RIE made us aware of a recent monograph in which Cronbach advocated an analysis which appears to accomplish the same result by a different procedure.

In a way it seems strange that an analysis which is widely used in agriculture, the split-plot analysis, should be so ignored in the social sciences. It may be because the agricultural application has typically been concerned with differences between means across classifications, rather than with relationships between continuous variables. And perhaps it may be because analysis of large numbers of classification variables has been cumbersome with the statistical libraries best known in the social sciences.

The analysis which was evolved studies relations at two levels, the classroom and the pupil, in successive steps, using a standard library program. In the first step, mean pupil characteristics and classroom process measures and their interactions are related with gain across classrooms. This is parallel to the usual analysis of classroom means. Following this, all remaining variance due to differences between classrooms is identified, which provides an error term for the between classroom analysis. At the next step, relations between individual pupil characteristics and classroom process with gain are studied. Since all classroom level effects have previously been partialled out, gains of individual pupils are independent of the information conveyed by class membership. The remaining variance is used as the error term for the within class analysis.

Exploration of the use of means for subgroups aggregated within classroom on the basis of pretest characteristics appeared to produce spurious results, so the two level analysis, class and pupil, was carried out using the data of individual pupils. Even if the procedure is used only for between class analyses, it seems to be preferable to the usual analysis of means since each classroom is weighted by the number of pupils represented in the analysis.

In interpreting the percentages of variance which are reported in the tables, it is important to remember that they were calculated on different bases from between class to within class. For relations between classes, the sum of squares for between classes was used as the base, whereas for the within class, the total sum of squares was used as the base. Since the between class sum of squares typically ranged from about 8 to 13 percent of the total sum of squares, the two sets of percentages are not comparable. The between class sum of squares was used as a base to make the between class percentages comparable with most of the studies in the field which have analyzed class means. To make them comparable to the within class percentages, they would need to be divided by constants ranging from 8 to 11, depending on the measure.

Usefulness of Within Class Analysis

A major goal of the project was to examine interactions of individual pupil characteristics with classroom process, with the expectation that

this would further clarify the nature of effective teaching. Significant interactions did appear between pupil characteristics and classroom process, within class. While they offered support for clarifying effective teaching, it was moderate since the variances accounted for were generally smaller than for the between class analyses.

The usefulness of the within class analysis appeared to be greater for pupil characteristics, however, where relations with gain were typically as strong as for the between class analyses.

The procedure does appear to be worth further exploration. In particular, it would be appropriate to apply it to data in which the interaction of the teacher with individual pupils has been recorded. It may be that variability of the individual teacher from pupil to pupil within class is considerable. If so, the within class variance would be stronger. If not, that analysis is likely to produce results similar to these. But the question is one worth examination.

Further, the analysis is applicable to any problem in which study of relations at several levels is relevant: the study of differences between teachers, holding school effects constant; differences between schools, holding school system effects constant; differences between school systems, holding states or regions constant; examining relations between variables at any level, while holding constant the effects of differences or relationships at higher levels--all are feasible, as are extensions to more than two levels of analysis. For example, in the current discussion of how the Follow Through evaluation data should have been analyzed, (House, Glass, McLean, and Walker, 1978) it seems to us that it would have been desirable to analyze programs partialling effect of site, so that site as a contaminating source of variability would have been eliminated from the analysis.

The Use of Models to Represent Process-Product Relations

For each of the levels of the split-plot analysis, a two-stage process of analysis was followed. The first stage was a step-wise multiple regression which was used to identify a small number of variables for further analysis. The second stage consisted of testing models--one or more made up only of first order terms, and one or more which included second order terms--using SAS GLM. This two stage process was followed because GLM will not select variables on the basis of variance accounted for but will conveniently process classification or nominal variables (classrooms), and will provide the sum of squares for each variable in the analysis which would have been obtained if that variable had been the last variable entered into the regression equation. Since these sums of squares are not dependent on the order of entry, they are the ones which were used and reported.

Testing promising classroom behavior measures in models appears to be a procedure worth further use. It is often true that examination of the zero order correlations (such as those in Appendix A) would lead to rather different conclusions about effective teaching than would the same measures in a model. Partialling out the effects of the other variables

in the model clearly alters the picture. Where two variables are colinear (share variance accounted for), the overlap is recognized in the model, and only the unique variance is identified. The model also takes account of suppression effects (a variable accounts for more variance in the outcome when others are partialled), which may parallel the effects of experimental control in which masking effects are held constant.

Further, in the application of multiple regression in which two or more entry levels are used, a large number of variables is often present in the equation before second level terms are reached, so the extent to which the results depend on the presence of numbers of nonsignificant variables is unknown. This problem is minimized by screening down the variable set to a small number which hopefully accounts for much of the variance accounted for by the larger equation.

While it is true that even the small set of variables will be inflated to a degree in the variance accounted for if the original set was large, as it would be of the larger multiple regression equation, the use of models appears to be warranted. The use of restricted models and of models of several degrees of complexity seems to be a procedure which presents a clearer, and probably more accurate, picture of the relations between a set of independent variables and a dependent variable.

Pupil Characteristics and Process-Product Relations

Although the study of process-product relations between classrooms was not an objective of the study, discussion of those results seems appropriate since different process measures have been created which may produce somewhat different results, and since the analysis by way of models also produces somewhat different results. However, since the results for each grade level have been summarized earlier, only results of particular interest or results for which there are parallels or contrasts across grade levels will be discussed.

Pupil Characteristics and Gain

A first finding which applies to both first and fifth grades is that there was considerable variance in gain accounted for between classrooms by measures of pupil characteristics at the beginning of the year, despite the fact that sex, race, pretest, and either socioeconomic status or IQ had been partialled out of the gain measures previously. For fifth grade, there was a strong negative relation between Anxiety and gain in Reading. In the first grade, Draw-A-Man and Familiar Figures showed strong positive relations with gain in three of the five subtests. Since these variables were represented in the models in which the effects of classroom behavior were studied, the process-product relations are almost certainly different than they would have been for zero order correlations. These findings, along with a previous finding that classroom behavior tended to be relatively strongly related with pupil characteristics at the beginning of the year, underlines the need to partial out pretest pupil characteristics as completely as possible so as not to have them reflected in both the classroom behavior and the measure of gain, producing spurious relationships.

For the within class analyses, when the difference in base is recognized, considerable amounts of variance in gain are accounted for by pupil characteristics at both grade levels. At the first grade, pretest Alphabet, which was included as a possible reflection of home support for learning, was positively related with four of the five outcome measures. Draw-a-Man and Familiar Figures each was positively related with two of the outcomes. At the fifth grade, pretest Reading was positively related with two outcomes; Dependency was positively related with one, and Anxiety was negatively related with one.

It is perhaps worth a reminder that these relations were obtained after classroom level effects had been partialled out.

Process-Product Relationships

The process measures also accounted for meaningful amounts of variance. One of the surprising results was that a measure of Teacher and Pupil Positive Affect was related strongly and consistently negatively with gain measures at both grade levels. This appears to be contrary to one of educators' most strongly held beliefs. While the classroom measure may reflect lack of task focus, the fact that other effects have been partialled out raises a real possibility that positive affect can be used in ways which hinder learning. It seems clear that we need to know more about the context and use of positive affect in the classroom.

A measure of Teacher Strong Control and Negative Affect was related strongly negatively with gain in two outcome measures at first grade, and Pupil Negative Affect and Disorder was related negatively with gain in Reading at the fifth grade. The contrast of these measures at these grade levels suggests that teacher negative behavior may be more harmful at first grade but that pupil negative affect and disorder may be more harmful at fifth grade, although this interpretation was not verified by a grade level interaction.

A measure of Pupil Interest in and Attention to Task was strongly positively related to Spelling gain at the fifth grade, but was strongly negatively related with gain in Copying at the first grade, and just short of significance for Numbers, also negative. Spelling was hypothesized to be a simple outcome measure, whereas Numbers was judged to be complex, and Copying would also have been a reasonable choice since it parallels an item from the Stanford-Binet. These findings suggest, then, that the relation of Interest-Attention to gain might depend on grade level, but that seems doubtful since other studies have found task involvement to be positively related with gain in the lower grades. Another possibility is that it depends on the complexity of the outcome measure, relating positively with gain in simple measures and negatively with complex ones. This would only follow if attention to task, as we measured it, was more easily observed when the pupil was engaged in a simple task, but that does seem possible. For example, if a pupil is staring into space, it is hard to judge whether he is daydreaming or processing information.

A measure of Mid to High Cognitive Level interaction, was positively related to Reading gain, and a measure of Narrow, One-Answer interaction was positively related with gain in Spelling and negatively, although nonsignificantly, with gain in Reading. These relationships will be discussed in the next section.

A final question is that of why nonlinear relations between classroom behaviors and pupil gain did not emerge in these analyses, in contrast to those which emerged in earlier analyses of these same data. One possibility may be that the paradigm process measures used in these analyses tended to be narrower and more specific than the traditionally derived factors which were used in the earlier analyses. Another possibility may be that pupil characteristics were controlled more extensively in these analyses than was true of the earlier ones. We do not understand why these changes should eliminate nonlinear relations, but they are the principal changes made from the earlier project to this one. It does seem clear that the different procedure for producing process scores has produced different results in other ways and the absence of nonlinear relations may be an additional consequence.

Differences in Cognitive Level of Outcomes

It was hypothesized that there would be differences in the process measures which would be related to gain as a function of the cognitive level or the complexity of the outcome measure. We can offer no proof that some outcomes were more complex than others. The classification of subtests was judgmental, based on inferences about the intellectual operations a child at that grade level would carry out to answer the test items. But in support of this procedure, the taxonomy for the cognitive domain was initially developed to classify test items.

The fifth grade verbal measures of Vocabulary and Reading were judged to be complex, and Spelling was judged to be a simple outcome measure. Arithmetic Concepts was judged to be a complex quantitative measure, but it had so little variance in common with classroom process as measured here that it had little to say about teaching effectiveness.

Gain in Spelling was positively related with the classroom process measures of Narrow, One-Answer interaction and Pupil Interest in and Attention to Task. Gain in Reading, a complex measure, was positively related with Mid to High Cognitive level interaction, but neither it nor Vocabulary was significantly related with either Narrow, One-Answer or Interest-Attention, the two variables which were related with Spelling gain.

For the first grade, Word Meaning was judged to be the simplest of the measures and Numbers was judged to be complex, although Copying and Matching might also have been selected as complex. The only process measure which might be taken as being related to cognitive level was Interest-Attention. It was negatively related with Copying, and approached a significant negative relation for Numbers, in contrast with the positive relation with gain in Spelling, at fifth grade. While the difference in

A final interaction suggested that classrooms with high Pupils Follow Routine supported gain for pupils who were internal in control. Apparently where pupils were on their own to follow preestablished procedures, those who were inner directed gained more than those who were not.

In summary, for the Between Class analyses, the significant interactions between mean pupil characteristics and classroom behavior were not numerous compared to the number of process measures which were significant, but they were relatively strong in terms of variance accounted for, ranging from 9 to 16 percent. Only one of the first order terms accounted for more than 16 percent of variance.

For the Within Class analyses, the percentages of variance were smaller, but as discussed earlier, were calculated on a different base. The Within Class percentages were calculated using the total sum of squares for the analysis, but the Between Class percentages used the Between Class sum of squares as the base. Since the Between Class sum of squares was from 7 to 13 percent of the total sum of squares, to be comparable, the within percentages would need to be multiplied by constants ranging from 8 to 11, depending on the outcome measure. If that is done, the variances accounted for Within Class range from under 3 to nearly 15 percent of variance accounted for. These were less than most of the Between Class interactions, but they still seem worth consideration.

For first grade there were two of the Within Class interactions which were significant. For Word Meaning, in classrooms where Differentiated Activities were low, pupils who were reflective gained more than pupils who were impulsive. Perhaps there is a parallel between this interaction and the fifth grade Between Class interaction which showed that classrooms with highly Differentiated Activities supported gain for low motivated pupils. Since the reflective pupil was willing to spend time considering the details of highly similar figures, perhaps he parallels in some ways the pupil who is motivated.

The other interaction within classrooms at the first grade level was for Listening. It indicated that in classrooms where Teacher Strong Control and Negative Affect was high, white pupils learned more than black pupils; where there was little Teacher Strong Control, black pupils learned more than white. At the risk of oversimplification, black pupils seemed to be more hindered by Teacher Strong Control than white pupils.

For the fifth grade Within Class analyses there were two interactions for Spelling. They indicated that in classrooms where Pupils Follow Routine was low, dependent pupils gained more than independent, and high anxious pupils gained more than low anxious. Perhaps in the absence of a set of limits which restrain pupil activity, anxious pupils gained because they were internally driven, and dependent pupils gained because it was expected.

A Concluding Comment

These results offer support for Cronbach's point that analyzing data in a way which identifies both between and within classrooms effects is functional and important.

They suggest that the nature of the teaching which is most effective depends on both the characteristics of the pupils being taught and the complexity of the outcome measure. Findings such as these are not likely to simplify the task of the teacher or tercher educator. Would we then be better off to ignore the research literature or to attend only its simpler findings? One way to approach the question of whether it is worthwhile to attempt to follow the complexities of interactive relationships in contrast to simple, linear ones is to ask how much variance each set of variables accounts for. Although it is surely not a definitive answer, for Reading and Vocabulary, for the fifth grade between classes analysis, each second order (interactive) variable accounted for as much or more variance as first order (simple,linear) terms did, on the average, and they did so when all other sources of variance in the model were first partialled out. Apparently, complex relations represent a significant fraction of the variance we can account for in the real world.

Again, it should be remembered that a large number of variables was used in relation to the number of degrees of freedom for the class level analysis, and that uncertainty in meaning of results follows. This uncertainty along with the complexity of the results, urge caution in drawing recommendations for classroom procedures. The results do argue for reanalyses of existing data sets using both between and within class analyses, where pupil characteristics have been measured.

It is also clear that noncognitive characteristics of pupils such as anxiety, internality, reflectivity and motivation in and of themselves account for significant and practically important proportions of variance in achievement gain. Knowledge of these relations adds to understanding of the learning process, and probably changes process-product relationships.

Although this analysis seems clearly to be a necessary extension of methodology, it is equally clear that at least for now, it will not provide a simple, readily useable conception of effective teaching which takes into account differences in the pupils being taught. Rather, it adds to the complexity. But it also suggests that we ignore that complexity at some cost.

References

Amidon, E. J., & Hough, J. B. Interaction Analysis: theory, research and application. Reading, Mass.: Addison-Wesley, 1967.

Barr, A. J., Goodnight, J. H., Sall, J. P., & Helwig, J. T. A user's guide to SAS 76. Raleigh, N. C.: SAS Institute, Inc., 1976.

Bereiter, C. Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. Taxonomy of educational objectives Handbook I: cognitive domain. New York: McKay, 1956.

Brophy, J. E., & Evertson, C. M. The Texas teacher effectiveness project: Presentation of non-linear relationships and summary discussion, Report No. 74-6. Austin: Research and Development Center, University of Texas, 1974.

Brown, B. B. The experimental mind in education. New York: Harper & Row, 1968.

Campbell, D. T. & Erlebacher, A. How regression artifacts can mistakenly make compensatory education look harmful. In J. Helmuth (Ed.) Disadvantaged Child (V.3), Chpt. 10. New York: Brunner/Marzel, 1970.

Castaneda, A., McCandless, B. R., & Palermo, D. S. The children's form of the manifest anxiety scale. Child Development, 1956, 27, 317-323.

Castaneda, A., Palermo, D. S., & McCandless, B. R. Complex learning as a function of anxiety in children and task difficulty. Child Development, 1956, 27, 327-332.

Cohen, J. & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. New York: Wiley, 1975.

Cronbach, L. J. Research on classrooms and schools: Formulation of questions, design, and analysis. Stanford Evaluation Consortium, Stanford Univ., Stanford, Calif., 1976. ERIC Document No. ED 135 801.

Cronbach, L. J. & Snow, R. E. Aptitudes and instructional methods. New York: Halsted Press, 1977.

Dogan, M. & Rokkan, S. (Eds.). Quantitative analysis in the social sciences. Cambridge, Mass.: MIT Press, 1969.

Duncan, O. D. A socioeconomic index for all occupations. In A. J. Reiss (Ed.), Occupation and Social Status. New York: Free Press of Glencoe, 1961.

Dunkin, M. J., & Biddle, B. J. The study of teaching. New York: Holt, Rinehart, & Winston, Inc., 1974.

Flanders, N. A. Teacher influence, pupil attitudes, and achievement. Cooperative Research Monograph No. 12, OE-25040, U. S. Dept. of HEW, 1964.

Flanders, N. A. Analyzing teacher behavior. Reading, Mass.: Addison-Wesley, 1970.

Flanders, N. A., Anderson, J. P., and Amidon, E. J. Measuring dependence proneness in the classroom. Educational & Psychological Measurement, 1961, 21, 575-587.

Fleishman, E. A. Leadership climate, human relations training, and supervisory behavior. Personal Psychology, 1953, 6, 205-222.

Fowler, D. Relation of teacher personality characteristics and attitudes to teacher-pupil rapport and emotional climate in the elementary classroom. Unpublished doctoral dissertation, University of South Carolina, 1962.

Freeman, F. S. The Lorge-Thorndike intelligence test. In Buros, O. K. (Ed.), The Fifth Mental Measurement Yearbook. Highland Park, New Jersey: Gryphon Press, 1959.

Frymier, J. R. Development of validation of a motivation index: a sixth report. Columbus, Ohio: Ohio State Univeristy, 1965 (mimeo).

Glass, G. V., & Stanley, J. C. Statistical methods in education and psychology. Englewood Cliffs, N. J.: Prentice Hall, 1970.

Hildreth, G. H., Griffiths, N. L., & McGauvran, J. E. Metropolitan Readiness Tests. New York: Harcourt, Brace and World, Inc., 1969.

Hines, V. A. Personal communication, n.d.

House, E. R., Glass, G. V., McLean, L. D., & Walker, D. F. No Simple Answer: Critique of the Follow Through Evaluation. Harvard Educational Review, May 1978, 48(2), 128-160.

Kagan, J. Reflection-impulsivity and reading ability in primary grade children. Child Development, 1965, 36, 609-28.

Kagan, J., Pearson, L., & Welch, L. Conceptual impulsivity and inductive reasoning. Child Development, 1966, 37, 583-94.

Lindquist, E. F., & Hieronymus, A. N. Iowa Tests of Basic Skills. New York: Houghton Mifflin, 1956.

McNemar, Q. On so-called test bias. American Psychologist, 1975, 30, 848-851.

Medley, D. M. Teacher competence and teacher effectiveness: A review of process-product research. Washington, D. C.: American Association of Colleges for Teacher Education, 1977.

Medley, D. M., & Mitzel, H. E. A technique for measuring classroom behavior. Journal of Educational Psychology, 1958, 49, 86-92.

Medley, D. M. & Mitzel, H. E. Some behavioral correlates of teacher effectiveness. Journal of Educational Psychology, 1959, 50, 239-246.

Medley, D. M. & Mitzel, H. E. A revision of the Observation Schedule & Record, (mimeo), n.d.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner K., & Bent, D. H. Statistical package for the social Sciences. New York: McGraw-Hill, 1975.

Ober, R. L., Wood, S. E., & Roberts, A. An instructional manual for the Reciprocal Category System. Based on a paper read at the annual meeting of the American Educational Research Association, Chicago, Illinois, February 8-10, 1968.

Peckham, P. D., Glass, G. V., & Hopkins, K. D. The experimental unit in statistical analysis. Journal of Special Education, 1969, 3, 337-349.

Raths, James. The appropriate experimental unit. Educational Leadership, 1967, 25, 263-266.

Rosenshine, B. Teaching behaviors and student achievement. London: International Association for the Evaluation of Educational Achievement, 1971.

Rosenshine, B. V., & Berliner, D. C. Academic engaged time. British Journal of Teacher Education, 1978, 4, 3-16.

Rosenshine, B., & Furst, N. The use of direct observation to study teaching. In R. M. W. Travers, (Ed.), Second handbook of research on teaching. Chicago: Rand McNally, 1973, 122-183.

Rotter, J. Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 1966, 80(1).

Rowe, M. B. Relation of wait-time and rewards to the development of language, logic, and fate control: Part II-rewards. Journal of Research in Science Teaching, 1974, 11(4), 291-308.

Sanders, N. M. Classroom questions: What kind? New York: Harper & Row, 1966.

Soar, R. S. An integrative approach to classroom learning, Final Report. ERIC Document No. ED 033 749.

Soar, R. S. Optimum teacher-pupil interaction for pupil growth. Educational Leadership Research Supplement, 1968, 1, 275-280.

Soar, R. S. Follow Through classroom process measurement and pupil growth (1970-71) final report. Gainesville, Florida: Institute for Development of Human Resources, University of Florida, 1973. ERIC Document No. ED 106 297.

Soar, R. S., & Soar, R. M. An empirical analysis of selected Follow Through programs: An example of a process approach to evaluation. In I. J. Gordon (Ed.), Early childhood education. Chicago: National Society for the Study of Education, 1972.

Soar, R. S., & Soar, R. M. Classroom behavior, pupil characteristics, and pupil growth for the school year and for the summer. Gainesville, Florida: Institute for Development of Human Resources, University of Florida, 1973. JSAS Catalog of Selected Documents in Psychology, 1975, 5, 200. (Ms. No. 873).

Soar, R. S., & Soar, R. M. Emotional climate and teacher management: A paradigm and some results. In H. J. Walberg & P. L. Peterson (Eds.), Conceptions of teaching. Chicago: National Society for the Study of Education and McCutchan Publishing Corp., in press.

Stallings, J. Implementation and child effects of teaching practices in Follow Through classrooms. Monographs of the Society for Research in Child Development, 1975, 40(7 and 8).

Warner, W. L., Meeker, M. & Eells, K. Social class in America. Chicago: Science Associates, 1949.

Wiley, D. E., & Bock, R. D. Quasi-experimentation in educational settings: comment. The School Review, 1967, Winter.

Appendix A
Table A-1

Correlation of Paradigm Factors with
Regressed Gain Class Means for Fifth Grade

Factor	Vocab[1]	Reading[2]	Spelling[1]	Arith Conc[3]
P6 T Gentle to Moderate Control	.08	-.13	.07	.12
P8 Pupils Follow Routine	-.01	-.07	.20	.10
P9 Low Cognitive Level	-.15	.12	-.09	-.11
P10 Mid to High Cognitive Level	-.02	.22	.14	.04
P11 T Central and Directed	-.21	.00	-.03	-.07
P12 Differentiated Activities	.25*	.02	.01	.06
P15 P Interest-Attention	.19	.27*	.40**	.27*
P16 T Guided Discovery	.09	-.15	.07	.12
P17 Narrow, One-Answer	-.07	.01	.22	.03
P18 Pupils Guess	.06	-.03	-.14	-.22
P19 T Strong Control & Neg. Aff.	-.00	-.07	-.17	-.13
P20 P Negative Affect & Disorder	.04	-.29*	-.36**	-.09
P21 T & P Positive Affect	.03	-.16	-.03	.09

[1] N = 59 * $p < .05$ ** $p < .01$

[2] N = 52

[3] N = 53

Appendix A
Table A-2

Correlation of Paradigm Factors with
Regressed Gain Class-Means for First Grade[1]

Factor	Word Meaning	Listening	Matching	Numbers	Copying
P6 T Gentle to Moderate Control	-.30	-.31	-.49*	-.07	-.01
P8 Pupils Follow Routine	-.18	.38	.24	.10	.18
P9 Low Cognitive Level	-.06	.23	.02	-.17	-.11
P10 Mid to High Cognitive Level	-.15	.03	.25	.24	-.18
P11 T Central and Directed	-.16	-.19	.03	.33	-.24
P12 Differentiated Activities	-.25	.00	-.01	-.33	.15
P15 P Interest-Attention	-.20	.24	.22	-.29	-.14
P16 T Guided Discovery	-.27	.10	.33	.12	-.14
P17 Narrow, One-Answer	.16	.12	.26	.06	-.08
P18 Pupils Guess	-.19	.10	-.20	-.24	.15
P19 T Strong Control & Neg. Aff.	-.12	-.61**	-.43*	.15	-.36
P20 P Negative Affect & Disorder	-.18	-.45*	-.43*	.02	.18
P21 T & P Positive Affect	-.33	.22	-.00	-.16	.13

[1] N = 22 *p < .05 **p < .01